BMJ Open

PROTOCOL FOR A SYSTEMATIC REVIEW OF MATERNITY PATIENT-REPORTED OUTCOME MEASURES (PROMs) AND PATIENT-REPORTED EXPERIENCE MEASURES (PREMs): SUPPORTING THE DEVELOPMENT OF A WOMAN-CENTRED INSTRUMENT DATABASE

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-058952
Article Type:	Protocol
Date Submitted by the Author:	03-Nov-2021
Complete List of Authors:	Bull, Claudia; Monash University, Monash Centre for Health Research and Implementation; Monash University, School of Public Health and Preventive Medicine Teede, Helena; Monash University, Monash Centre for Health Research and Implementation; Monash University, Monash Partners Academic Health Science Centre Carrandi, Lane; Monash University, Monash Centre for Health Research and Implementation; Monash University, School of Public Health and Preventive Medicine Rigney, Azure; Maternity Choices Australia Cusack, Sally; Maternity Choices Australia Callander, Emily; Monash University, Monash Centre for Health Research and Implementation; Monash University, School of Public Health and Preventive Medicine
Keywords:	HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, OBSTETRICS

PROTOCOL FOR A SYSTEMATIC REVIEW OF MATERNITY PATIENT-REPORTED OUTCOME MEASURES (PROMs) AND PATIENT-REPORTED EXPERIENCE MEASURES (PREMs): SUPPORTING THE DEVELOPMENT OF A WOMAN-CENTRED INSTRUMENT DATABASE

Claudia Bull[1,2], Helena Teede[1,3], Lane Carrandi[1,2], Azure Rigney[4], Sally Cusack[4], Emily J Callander[1,2]

[1]Monash Centre for Health Research and Implementation, Monash University, 43-51 Kanooka Grove, Clayton, VIC, 3168, AUSTRALIA

[2]School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, VIC, 3004, AUSTRALIA

[3]Monash Partners Academic Health Science Centre, 43-51 Kanooka Grove, Clayton, VIC, 3168, AUSTRALIA

[4]Maternity Choices Australia, Springwood, Queensland, 4127, AUSTRALIA

Corresponding author:

Miss Claudia Bull

School of Public Health and Preventive Medicine, Monash University

Level 4, 553 St Kilda Road, Melbourne, VIC, 3004, AUSTRALIA

E: claudia.bull@monash.edu

Ph: +61 (03) 9903 0723

Word count: 3,259

ABSTRACT

Introduction: Woman-centred care is the right of every woman receiving maternity care, irrespective of where care is being received and who is providing care. This protocol describes a planned systematic review that will identify, describe, and critically appraise the psychometric properties of maternity patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs). The woman-centricity of PROM and PREM development and content validation will also be assessed. This information will be used to develop a maternity PROMs and PREMs database to support service and system performance measurement, and value-based maternity care initiatives.

Methods and analysis: This study will be guided by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Guideline for Systematic Reviews of Outcome Measurement Instruments. Studies identified via MEDLINE, CINAHL Plus, PsycINFO, and EMBASE describing the development, content validation, and/or psychometric evaluation of PROMs and PREMs designed to evaluate maternity care provided throughout pregnancy, childbirth and postnatal periods will be considered if published from 2010 onwards, in English, and available in full-text. The COSMIN risk of bias checklist will be used to evaluate the quality of studies reporting on the development, content validation, and/or psychometric evaluation of PROMs and PREMs. COSMIN criteria for good content validity will be used to assess the woman-centricity of PROM and PREM development and content validation studies. COSMIN standards of good psychometric properties will be used to evaluate the validity and reliability of the identified instruments.

Ethics and dissemination: Ethical permission for this research is not required. The findings of this research will be submitted for publication in an international, peer-reviewed journal. Abstracts for national and international conference presentations will also be submitted. The proposed maternity PROMs and PREMs database will be freely accessible online, and developed with consumer input to ensure its usefulness to a range of maternity care stakeholders.

PROSPERO registration: ID 288854

Keywords: Patient-reported outcome measure (PROM); Patient-reported experience measure (PREM); Woman-centred care; Survey development; Psychometric evaluation; Validity; Reliability

STRENGTHS AND LIMITATIONS OF THIS STUDY

- Employing the COSMIN guidance at all stages provides a transparent, uniform and robust approach to the conduct of this systematic review.
- Compiling evidence on the woman-centricity of maternity PROM and PREM development and content validation is yet to be evidenced in the peer-reviewed literature, and aims to support performance measurement and value-based assessment that is meaningful to women.
- Developing a publicly available database of maternity PROMs and PREMs aims to promote best practice instrument selection and implementation to support the measurement of services and systems, and contribute to operationalising value-based health care.
- A potential limitation of this review is using COSMIN guidance (developed for PROMs) to evaluate the development, content validation and psychometric evaluation of PREMs.

INTRODUCTION

The concept of woman-centred care (WCC) is underpinned by the principles of choice, control, continuity of carer, and a woman's right to self-determination.(1-3) WCC is typically associated with midwifery practice,(4) but this misrepresents the reality that receiving WCC is the right of every woman, irrespective of where or by whom she receives care. Coupled with a "risk avoidance" obstetric culture and increasing rates of intervention at birth (particularly in high-income countries),(3, 5, 6) women's values and preferences for aspects of care beyond a successful live birth (e.g., desire for a natural birth) are often a secondary consideration. This has subsequently challenged the implementation of value-based maternity care, where consumer perspectives are at the centre of outcome measurement.

Value-based healthcare (VBHC) is the purported goal of every health system. At its core, VBHC aims to improve patient health outcomes relative to the cost of achieving those improvements.(7) However, VBHC frameworks that exist on this principle alone have been called into question as they oversimplify the complex construct of 'value';(8) particularly what value means to patients in different circumstances.(9) Indeed, in the context of maternity care, women value a diverse array of factors including care continuity, equitability, promoting normal reproductive processes, choosing where they give birth, being treated respectfully, emotional support, and transparent communication.(10-13) Consequently, value-based maternity care represents far more than a successful live birth.

A key challenge to operationalising value-based maternity care is that values and preferences are heterogeneous (14). Consequently, a one-size-fits-all approach to value-based assessment is likely to be inappropriate to support tailored WCC practices across all health services. It is likely that different measures will be required in different settings to ensure that what matters to the women accessing services is being captured.

Patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs) are designed to measure and evaluate service and system performance from the consumer's perspective.

PROMs measure an individual's health and wellbeing.(15, 16) They can capture a wide-range of outcomes, largely related to physical, social, and/or psychological functioning.(15, 17, 18) A recent review of condition-specific PROMs used during pregnancy and childbirth revealed that outcomes such as nausea, vomiting, tiredness, emotional state, anxiety, depression and pain are commonly captured.(19) PREMs differ in that they are designed to measure an individual's experience of receiving care; namely, what happened during a care encounter and their perception of how it happened.(20) A recent concept analysis identified several constructs commonly captured in relation to women's experiences of maternity care. These include organisational aspects of care such as access and referral to maternity services, continuity of care, privacy and care costs; and interpersonal aspects of care such as information sharing, informed choice, emotional support, being treated with respect and dignity, and having confidence in the knowledge and ability of maternity care providers.(21) Both types of instrument contribute to measuring value by describing the outcomes and experiences of women accessing maternity care services. However, this is only achieved if the content of PROMs and PREMs aligns with what is viewed as important and relevant to women. Thus, woman-centric instrument development and content validation is crucial to supporting meaningful value-based measurement.(22)

We intend to develop a database hosting a repository of PROMS and PREMs to support the use of these instruments in practice as a part of achieving value-based maternity care. This protocol describes the systematic process that will be undertaken to firstly, identify and describe maternity care PROMs and PREMs published in the peer-reviewed literature, and secondly, critically appraise and summarise the psychometric properties of the identified instruments. Particular emphasis will be placed on the woman-centeredness of PROM and PREM development. The database will subsequently summarise this information in a user-friendly format suitable for a range of maternity care stakeholders.

METHODS AND ANALYSIS

This study will be guided by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) *Guideline for Systematic Reviews of Outcome Measurement Instruments*.(23)

COSMIN stipulates a 10-step process for performing a systematic review of PROMs (which will be extended to PREMs for the purposes of this research). Steps 1-4 pertain to conducting the literature search; steps 5-7 pertain to evaluating an instruments' psychometric properties; step 8 pertains to evaluating the interpretability and feasibility of implementing instruments; and steps 9-10 pertain to writing the review discussion. This protocol will detail the processes we intend to undertake for steps 1-8.

Step 1: Formulate the aim of the review

The aim of this review is two-fold. First, to identify and describe maternity care PROMs and PREMs published in the peer-reviewed literature. Second, to critically appraise and summarise the psychometric properties of the identified instruments, with particular emphasis on assessing the woman-centredness of instrument development.

Step 2: Formulate the eligibility criteria

Studies were included in the review if:

- Published from 2010 onwards, representing contemporary instruments (however, if articles refer to developmental and psychometric evaluation evidence pre-dating 2010, we will include these to provide a holistic representation of instrument quality)
- Published in English
- Available in full-text
- Described the development, content validation, and/or psychometric evaluation of a PREM designed to evaluate maternity care provided throughout pregnancy, childbirth and postnatal periods
- Described the development, content validation, and/or psychometric evaluation of a PROM designed to report on pregnant or postnatal woman's health condition related to physical, emotional and social function, signs or symptoms

Studies were excluded if:

- Published before 2010 (except as specified above)
- Published in languages other than English
- Not available in full-text
- Described the development, content validation, and/or psychometric evaluation of a PROM or PREM designed for a non-maternity care context (e.g., generic, inpatient, primary care) but used in a maternity care context
- Described the development, content validation, and/or psychometric evaluation of PREM designed to evaluate maternity care provided preconception (e.g., IVF, contraception)
- Described the development, content validation, and/or psychometric evaluation of a PROM not specific to the health condition of a pregnant or postnatal woman (e.g., PROMs related to abortion, lower back pain unrelated to pregnancy, prenatal genetic testing)
- Included a PROM or PREM as an outcome instrument (e.g., use of a maternity PROM in a randomised controlled trial), but were not contributing to the instruments' development, content validation, and/or psychometric evaluation
- Described PROMs or PREMs designed for care providers, children, or proxies (e.g., partner or carer completes the PROM or PREM on behalf of the woman)
- Described the development, content validation, and/or psychometric evaluation of a Health-related quality of life measure (HRQoL) or patient satisfaction measure
- Presented literature reviews, meta-reviews, protocols, theses or quality improvement activities

We specifically delineate PROMs from HRQoL measures. HRQoL measures are preference-based instruments despite often being referred to interchangeably as PROMs.(24-27) While PROMs and HRQoL measures may capture similar constructs,(28) they differ in how they are used and scored. PROMs are used to assess changes in health outcomes over time, and are scored on an item-by-item or domain/dimension basis. Conversely, HRQoL measures are used for the purposes of cost-utility analysis, where an individuals' quality of life as it relates to their health state is scored as one of a finite number of health states relative to a utility index.(29) Thus, HRQoL measures were designed to

estimate quality-adjusted life years and are used in the context of economic evaluation.(28) As such, they will not be included in this review.

We also specifically delineate PREMs from patient satisfaction measures, despite being referred to synonymously throughout the literature.(30) PREMs ask individuals to *report* on their experiences of care, where satisfaction measures ask individuals to *evaluate* their experiences. While report style questions aim to be objective,(15) evaluative questions are more likely to reflect an individuals' expectations, attitudes and desire to appear socially desirable, and are thus influenced by attributes peripheral to their care experience.(31-33) Additionally, where PREMs typically use frequency-based response scales (e.g., on a scale of never to always),(34-36) patient satisfaction measures tend to use agreement-based response scales (e.g., on a scale of strongly disagree to strongly agree),(37-39) which are more prone to acquiescence bias and straightlining.(40, 41) Thus given the differences between these instruments, patient satisfaction measures will not be included in this review.

Step 3: Perform a literature search

We will search the following electronic databases: MEDLINE (via Ovid), CINAHL Plus (via EBSCOhost), PsycINFO (via Ovid), and EMBASE (via Elsevier). Our search terms will include the following concepts: (i) maternity care and maternal health services; (ii) PROMs; (iii) PREMs; and (iv) measurement properties. We will employ the search terms developed by COSMIN relative to studies on measurement properties.(42) These are available for each of MEDLINE, CINAHL, PsycINFO and EMABSE. An example of our proposed MEDLINE search strategy is available in **Supplementary file 1**. Searches will be limited to only studies published in English and available in full-text.

Step 4: Select abstracts and full-text articles

After being exported from electronic databases, all search results will be imported into Covidence.(43) Two reviewers will independently review all titles abstracts to determine which articles warrant full-text retrieval and review. Full text review will also be undertaken by two independent reviewers. Discrepancies at all stages will be addressed through reviewer consultation

and consensus, and if needed, engagement of a third reviewer. The reference lists of all included papers will be hand-searched for other potentially relevant studies.(23)

Steps 5-7: Evaluating the measurement properties of the included PROMs and PREMs

Data extraction – Characteristics of the included PROMs and PREMs

The following data will be extracted from included studies: (i) PROM/ PREM name; (ii) construct(s)/ domain(s) captured; (iii) target population and setting; (iv) mode of administration (e.g., online, postal), and administration time during perinatal care (e.g. antenatal, postnatal); (v) recall period; (vi) number of items; (vii) response options; and (viii) original language. This information will be used to describe the included PROMs and PREMs. Information will be extracted per study and grouped where there have been multiple studies conducted for one instrument. One reviewer will extract all data.

Evaluating the methodological quality of studies

Methodological quality will be evaluated in relation to maternity PROM and PREM development, content validation, and psychometric evaluation using the COSMIN Risk of Bias checklist.(44) This checklist details specific study design elements that are important when assessing the measurement properties of an instrument. Only study design elements relevant to the measurement properties presented in **Table 1** and reported in studies will be assessed for risk of bias. Criteria for study design elements are rated using a scale of 'very good', 'adequate', 'doubtful', inadequate', or 'n/a'. The lowest rating for any criteria will be used to describe the quality of the study underpinning that specific measurement property (i.e., worst score counts).(23) If multiple studies have been conducted to evidence a specific measurement property (e.g., three studies report on an instruments' internal consistency) and have provided variable results, the overall quality of the measurement property will be labelled 'unclear'.

One reviewer will first consult the *COSMIN database of systematic reviews of outcome measurement instruments*(45) to determine whether other researchers have already evaluated the risk of bias of the included studies. If available, existing ratings will be used. If not, one reviewer will determine the

measurement property(ies) that need to be assessed per study, and two reviewers will independently complete the Risk of Bias checklist for each individual study. We will use the Risk of Bias Microsoft Excel template developed by COSMIN to document each rater's scores.

Table 1: Elements of the COSMIN Risk of Bias checklist(44) for assessing study design relative to PROM and PREM development, content validation and psychometric evaluation studies

Measurement property	Number of criteria
Content validity	
PROM/ PREM development	35
Content validity	31
Internal structure	
Structural validity	4
Internal consistency	5
Cross-cultural validity/ measurement in variance	4
Remaining measurement properties	
Reliability	8
Measurement error	6
Criterion validity	3
Hypotheses testing for construct validity	7
Responsiveness	13

PROM = Patient-reported outcome measure; PREM = Patient-reported experience measure

Evaluating the content validity (woman-centredness) of PROM and PREM development

Content validity has been described as the most important measurement property of PROMs (and arguably, PREMs).(46) It represents the degree to which the content of an instrument is an adequate reflection of the phenomena being measured.(47) PROM and PREM items need to demonstrate appropriate relevance, comprehensiveness and comprehensibility to qualify as content valid.(46) This assessment should be made by 'experts' of the target phenomena. In the context of maternity care, the women receiving and experiencing care are the experts. COSMIN also provide supplemental criteria to support studies that have asked health professionals about the relevance and comprehensiveness of items.(46) However, as we are aiming to determine the woman-centricity of current PROMs and

PREMs, instruments that fail to demonstrate appropriate involvement of women in their development and content validation will be labelled as demonstrating 'inadequate' content validity.

COSMIN has developed a set of instructions specifically for evaluating the content validity of PROMs which will be used in this study (for both PROMs and PREMs). The first two steps involve evaluating the quality of studies reporting on instrument development and content validation; this forms part of the COSMIN Risk of Bias checklist described above. The third step involves rating each development and content validation study against nine criteria for good content validity (**Table 2**).(46) For each of relevance, comprehensiveness and comprehensibility, if ≥85% of an instruments' items fulfil the criteria, the study is deemed to have sufficient (+) evidence; if <85% of items fulfil the criteria, the study is deemed to have insufficient (—) evidence; and if there is inadequate information available or the study quality was inadequate (as identified through risk of bias assessment), the study is deemed to have indeterminate (?) evidence.(46) From this, we will assign an overall content validation score (+, —, ?) which will represent the woman-centeredness of PROM and PREM development. One reviewer will undertake the content validation assessment.

Table 2: Relevance, comprehensiveness and comprehensibility criteria for evaluating the content validity of maternity care instruments(46)

Criteria
Relevance
1. Are the included items relevant to maternity care?
2. Are the included items relevant to women?
3. Are the response options appropriate?
4. Is the recall period appropriate?
Comprehensiveness
5. Are all key concepts included?
Comprehensibility
6. Are the instrument instructions understood by women as intended?
7. Are items and response options understood by women as intended?
8. Are items appropriately worded?
9. Do the response options match the question?

Evaluating the sufficiency of measurement properties

Instruments will next be evaluated according to how well the reported measurement properties (e.g., structural validity) comply with standards of good psychometric properties (**Table 3**).(23) This will indicate whether a PROM or PREM can be considered valid and reliable. Validity is the extent to which an instrument measures what is purports to measure.(48, 49) Reliability is the extent to which participant responses to an instrument can be replicated in unchanging circumstances (consistency).(50) Reliability is also the extent to which an instrument is devoid of measurement error.(51, 52)

Using the COSMIN updated criteria for good measurement properties, psychometric properties will be rated as + (provides sufficient evidence), ─ (provides insufficient evidence), and ? (provides indeterminate evidence) (**Table 3)**. Red text denotes criteria added based on prominence in the literature relative to instrument development and psychometric evaluation. COSMIN's criteria of 'Hypothesis testing for construct validity' has been excluded from Table 3 as the context of maternity care in this study is too broad for the review team to appropriately generate hypotheses suitable for all potential instruments. If a PROM or PREM has several studies reporting on its' psychometric properties, each study will be evaluated individually (according to the reported psychometric properties), and an overall conclusion regarding the quality of the instrument will be provided for each psychometric quality. Any psychometric properties not assessed will be labelled as having 'no evidence'. One reviewer will undertake the good psychometric properties assessment.

Table 3: COSMIN updated criteria for good measurement properties(23)

Measurement property	Rating	Criteria
Structural validity	+	**Classical test theory (CTT)** • Confirmatory factor analysis (CFA): CFI or TLI (or comparable measure) >0.95, **OR** RMSEA <0.06, **OR** SMSR <0.08; **AND/OR** • Exploratory factor analysis (EFA) or Principal Components Analysis (PCA): KMO ≥0.70, **AND** Significant Bartlett's Test of Sphericity (p<0.05), **AND** dimensional (total) variance explained ≥50% or dimensional (total) variance explained <50% but justified by the authors(53, 54) **Item Response Theory (IRT)/ Rasch:** • No violation of unidimensionality: CFI or TLI (or comparable measure) >0.95, **OR** RMSEA <0.06, **OR** SMSR <0.08; **AND**

		• No violation of local independence: residual correlations among the items after controlling for the dominant factor <0.20 OR Q_3 fit statistics <0.37; **AND** • No violation of monotonicity: adequate looking graphs, OR item scalability >0.30; **AND** • Adequate model fit – IRT: χ^2 >0.01; Rasch: infit and outfit mean squares between ≥0.50 and ≤1.50, OR z-standardised values between >-2 and <2
	?	**CTT** • Not all information for + reported **IRT/ Rasch** • Model fit not reported
	–	**CTT** • Criteria for + not achieved **IRT/ Rasch** • Criteria for + not achieved
Internal consistency	+	Evidence of sufficient structural validity achieved (+ or ? for 'Structural validity'); **AND** Cronbach's alpha(s) ≥0.70 for each unidimensional scale or subscale
	?	Evidence of sufficient structural validity not achieved
	–	Evidence of sufficient structural validity achieved (+ or ? for 'Structural validity'); **AND** Cronbach's alpha(s) <0.70 for each unidimensional scale or subscale
Reliability	+	ICC or weighted Kappa ≥0.70
	?	ICC or weighted Kappa not reported
	–	ICC or weighted Kappa <0.70
Measurement error	+	SDC or LoA < MIC
	?	MIC not defined
	–	SDC or LoA > MIC
Cross-cultural validity/ measurement invariance	+	No importance differences found between group factors (such as age, gender, language) in multiple group factor analysis; **OR** No important DIF for group factors (McFadden's R^2 <0.02)
	?	No multiple group factor analysis performed; **OR** No DIF analysis performed
	–	Important differences between group factor analysis identified; **OR** Important differences in DIF analysis identified
Criterion validity	+	Correlation with gold standard instrument ≥0.70*; **OR** AUC ≥0.70
	?	Not all information for + reported
	–	Correlation with gold standard instrument <0.70*; **OR** AUC <0.70
Responsiveness	+	AUC ≥0.70
	?	AUC not reported
	–	AUC <0.70

*Correlation with a gold standard will only occur if a short-form instrument is being compared against its long-form counterpart; CTT = Classical Test Theory; CFA = Confirmatory Factor Analysis; CFI = Comparative Fit Index;/ TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SMSR = Standardised Root Mean Residuals; EFA = Exploratory Factor Analysis; CA = Principal Components Analysis; KMO = Kaiser-Meyer-Olkin; IRT = Item Response Theory; ICC = Intraclass Correlation Coefficient; SDC = Smallest Detectable Change; LoA = Limits of Agreement; MIC = Minimally Important Change; DIF = Differential Item Functioning; Area Under the Curve

Step 8: Describe the interpretability and feasibility of instrument implementation

Interpretability is the extent to which meaning can be derived from participant responses to an instrument or changes in responses.(55) This may include distinct patterns of responses amongst

subgroups of the population, trends in responses over time, and floor or ceiling effects. For the purposes of this review, we will extract and describe the following features of PROM and PREM interpretability: (i) distribution of responses in the study population and relevant subgroups; (ii) proportion of missing data for items; (iii) methods of handling missing data; (iv) evidence of floor and ceiling effects; and (v) minimally important changes (MIC) or minimally important differences (MID) in responses. Interpretability, whilst not considered a measurement property in and of itself, is important for understanding the real-world application and biases associated with implementing PROMs and PREMs.

Feasibility refers to the ease and convenience with which a PROM or PREM can be implemented and administered in a real-world context.(23) For the purposes of this review, we will extract and describe the following features of PROM and PREM feasibility: (i) available modes of administration; (ii) length of the instrument; (iii) estimated completion time; (iv) level of readability; (v) ease of response calculation; (vi) copyright; (vii) cost of using an instrument; (viii) equipment required for instrument administration; (ix) availability of instrument for application in different settings and languages; and (x) approvals required before instrument use. For the development of a maternity PROMs and PREMs database, this information will be critical for informing the real-world implementation of maternity PROMs and PREMs across health services and systems.

Patient and public involvement statement

The research team comprises members of Maternity Choices Australia, a national consumer advocacy organisation committed to the advancement of best-practice maternity care.(56) These women are consumer representatives and have been involved in the conceptualisation of the research and protocol development, recognising the importance of operationalising woman-centred care, and ensuring that maternity services are consumer informed. Importantly, they will aid the development of the Maternity PROMs and PREMs database, supporting its usability by a range of maternity care stakeholders. They will also help disseminate the Maternity PROMs and PREMs database through formal and informal engagement with key collaborative parties.

ETHICS AND DISSEMINATION

Ethical permission for this research is not required as the review will only use information from previously published research. The findings of this research will be submitted for publication in an international, peer-reviewed journal. Abstracts will also be submitted for national and international conference presentations.

Maternity PROMs and PREMs database

We intend for the maternity PROMs and PREMs database to be freely accessible online, and useful to all individuals involved in maternity health services and systems performance measurement, and value-based maternity care. The design of the database will be consumer informed to ensure that it is easy to understand, and provides information relevant to a range of maternity care stakeholders. The psychometric results (structural validity, internal consistency, reliability, measurement error, cross-cultural validity/ measurement invariance, criterion validity and responsiveness) for each instrument will be summarised according to whether criteria were met (+), indeterminate (?), or not met (—) when all evidence for a specific instrument in considered collectively. The woman-centricity of instrument development will be similarly summarised according to the COSMIN criteria for good content validity. In addition, the database will summarise descriptive information for each instrument (e.g., number of items, domains captured, country of development); summarise information regarding each instruments' feasibility of use (e.g., copyright and reuse considerations, available modes of administration, costs etc.); and provide links to the studies describing instruments. For PROMs or PREMs not freely available, we will also provide the appropriate contact information for the instruments' original author or licensing agent.

We anticipate that the database will be updated annually. A member of the research team will re-run the search strategies (updating search terms as needed) and undertake the processes described in this protocol. This will support the identification of new instruments or additional evidence of PROM and PREM psychometric evaluation over time, ensuring that the database is up-to-date and aligns with advancements in PROM and PREM methodologies and results.

CONCLUSION

Supporting the selection and implementation of rigorously developed and psychometrically evaluated maternity PROMs and PREMs will be critical to enhancing value-based maternity care efforts. Moreover, understanding the extent to which women were involved in the development and content validation of maternity PROMs and PREMs will evidence which instruments support meaningful service and system performance measurement to women. By developing a publicly available database from the findings of this review, we hope to contribute to initiating and furthering value-based maternity care initiatives globally.

REFERENCES

1. Leap N. Woman-centred or women-centred care: Does it matter? British Journal of Midwifery. 2009;17(1):12-6.

2. Snowden A, Martin C, Jomeen J, Hollins Martin C. Concurrent analysis of choice and control in childbirth. BMC Pregnancy Childbirth. 2011;11:40.

3. Brady S, Lee N, Gibbons K, Bogossian F. Woman-centred care: An integrative review of the empirical literature. Int J Nurs Stud. 2019;94:107-19.

4. Fontein-Kuipers Y, de Groot R, van Staa A. Woman-centered care 2.0: Bringing the concept into focus. Eur J Midwifery. 2018;2(May).

5. Akileswaran CP, Hutchison MS. Making Room at the Table for Obstetrics, Midwifery, and a Culture of Normalcy Within Maternity Care. Obstet Gynecol. 2016;128(1):176-80.

6. Maffi I, Gouilhers S. Conceiving of risk in childbirth: Obstetric discourses, medical management and cultural expectations in Switzerland and Jordan. Health Risk Soc. 2019;21(3-4):185-206.

7. Porter ME, Teisberg EO. Redefining health care: creating value-based competition on results. Brighton, MA: Harvard business press; 2006.

8. Perfetto EM. ISPOR's Initiative on US Value Assessment Frameworks: A Missed Opportunity for ISPOR and Patients. Value Health. 2018;21(2):169-70.

9. Perfetto EM, Oehrlein EM, Boutin M, Reid S, Gascho E. Value to Whom? The Patient Voice in the Value Discussion. Value Health. 2017;20(2):286-91.

10. De Jonge A, Downe S, Page L, Devane D, Lindgren H, Klinkert J, et al. Value based maternal and newborn care requires alignment of adequate resources with high value activities. BMC Pregnancy Childbirth. 2019;19(1):428.

11. Fawsitt CG, Bourke J, Greene RA, McElroy B, Krucien N, Murphy R, et al. What do women want? Valuing women's preferences and estimating demand for alternative models of maternity care using a discrete choice experiment. Health Policy. 2017;121(11):1154-60.

12. Fawsitt CG, Bourke J, Lutomski JE, Meaney S, McElroy B, Murphy R, et al. What women want: Exploring pregnant women's preferences for alternative models of maternity care. Health Policy. 2017;121(1):66-74.

13. Renfrew MJ, McFadden A, Bastos MH, Campbell J, Channon AA, Cheung NF, et al. Midwifery and quality care: Findings from a new evidence-informed framework for maternal and newborn care. Lancet. 2014;384(9948):1129-45.

14. Mooney G. Is it not time for health economists to rethink equity and access? Health economics, policy and law. 2009;4(2):209-21.

15. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. Bja Educ. 2017;17(4):137-44.

16. Black N, Burke L, Forrest CB, Sieberer UH, Ahmed S, Valderas JM, et al. Patient-reported outcomes: pathways to better health, better services, and better societies. Qual Life Res. 2016;25(5):1103-12.

17. Williams K, Sansoni J, Morris D, Grootemaat P, Thompson C. Patient-reported outcome measures: Literature review. Sydney: ACSQHC; 2016 November 2016.

18. Centre for Health Services and Policy Research. What are PROMs? Vancouver: The University of British Columbia; [cited 2021 June]. Available from: https://patientreportedoutcomes.ca/what-are-pros/proms/.

19. Dickinson F, McCauley M, Smith H, van den Broek N. Patient reported outcome measures for use in pregnancy and childbirth: a systematic review. Bmc Pregnancy Childb. 2019;19.

20. Bull C, Byrnes J, Hettiarachchi R, Downes M. A systematic review of the validity and reliability of patient-reported experience measures. Health Serv Res. 2019;54(5):1023-35.

21. Beecher C, Devane D, White M, Greene R, Dowling M. Women's experiences of their maternity care: A principle- based concept analysis. Women Birth. 2020;33(5):419-25.

22. Terwee CB, Prinsen CAC, Chiarotto A, de Vet HCW, Bouter LM, Alonso J, et al. COSMIN methodology for assessing the content validity and PROMs: User manual. Amsterdam: COSMIN; 2018.

23. Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, de Vet HCW, et al. COSMIN methodology for systematic reivews of patient-reported outcome measures (PROMs). Amsterdam: COSMIN; 2018.

24. Jonsson A, Orwelius L, Dahlstrom U, Kristenson M. Evaluation of the usefulness of EQ-5D as a patient-reported outcome measure using the Paretian classification of health change among patients with chronic heart failure. J Patient Rep Outcomes. 2020;4(1):50.

25. Benson T, Williams DH, Potts HW. Performance of EQ-5D, howRu and Oxford hip & knee scores in assessing the outcome of hip and knee replacements. BMC Health Serv Res. 2016;16(1):512.

26. NHS Digital. Patient Reported Outcome Measures (PROMs) in England - 2011-2012, Special topic, EQ-5D™ Index analysis UK: NHS; 2014 [updated 12 October 2018; cited 2021 June]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms/patient-reported-outcome-measures-proms-in-england-2011-2012-special-topic-eq-5d-index-analysis.

27. Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Serv Insights. 2013;6:61-8.

28. Feeny DH, Eckstrom E, Whitlock EP, Perdue LA. Patient-Reported Outcomes, Health-Related Quality of Life, and Function: An Overview of Measurement Properties. A Primer for Systematic Reviewers on the Measurement of Functional Status and Health-Related Quality of Life in Older Adults. AHRQ Methods for Effective Health Care. Rockville (MD): AHRQ; 2013.

29. Whitehurst DG, Noonan VK, Dvorak MF, Bryan S. A review of preference-based health-related quality of life questionnaires in spinal cord injury research. Spinal Cord. 2012;50(9):646-54.

30. Bull C. Patient satisfaction and patient experience are not interchangeable concepts. Int J Qual Health Care. 2021;33(1).

31. Williams B, Coyle J, Healy D. The meaning of patient satisfaction: an explanation of high reported levels. Soc Sci Med. 1998;47(9):1351-9.

32. Noyes RW, Levy MI, Chase CL, Udry JR. Expectation fulfillment as a measure of patient satisfaction. American Journal of Obstetrics and Gynecology. 1974;118(6):809-14.

33. Williams B. Patient satisfaction: a valid concept? Soc Sci Med. 1994;38(4):509-16.

34. Centers for Medicare & Medicaid Services. HCAHPS Hospital Consumer Assessment of Healthcare Providers and Systems Baltimore, MD: CMS; 2021 [updated 2 February 2021; cited 2021 June]. Available from: https://hcahpsonline.org/en/survey-instruments/.

35. U.S. Centers for Medicare & Medicaid Services. Emergency Department CAHPS (ED CAHPS) Baltimore, MD: CMS.gov; 2020 [updated 11 May 2020; cited 2021 June]. Available from: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/CAHPS/ED.

36. Care Quality Commission. Adult Inpatient Survey 2020 UK: NHS; 2020 [cited 2021 June]. Available from: https://nhssurveys.org/surveys/survey/02-adults-inpatients/.

37. Grogan S, Conner M, Willits D, Norman P. Development of a questionnaire to measure patients' satisfaction with general practitioners' services. Br J Gen Pract. 1995;45(399):525-9.

38. Kinnersley P, Stott N, Peters T, Harvey I, Hackett P. A comparison of methods for measuring patient satisfaction with consultations in primary care. Fam Pract. 1996;13(1):41-51.

39. Hojat M, Louis DZ, Maxwell K, Markham FW, Wender RC, Gonnella JS. A brief instrument to measure patients' overall satisfaction with primary care physicians. Fam Med. 2011;43(6):412-7.

40. Lavrakas PJ. Acquiescence Response Bias. In: Lavrakas PJ, editor. Encyclopedia of Survey Research Methods. Thousand Oaks, CA: Sage Publications; 2008.

41. Dykema J, Schaeffe NC, Garbarski D, Assad N, Blixt S. Towards a reconsideration of the use of agree-disagree questions in measuring subjective evaluations. Res Social Adm Pharm. 2021.

42. Terwee CB, Jansma EP, Riphagen, II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009;18(8):1115-23.

43. Covidence.org. Covidence Melbourne: Covidence.org; 2021 [cited 2021 September]. Available from: https://www.covidence.org/.

44. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018;27(5):1171-9.

45. COSMIN. COSMIN database of systematic reviews of outcome measurement instruments Amsterdam: University Library Vrje Universiteit Amsterdam; 2019 [updated Septmeber 2019; cited 2021 September]. Available from: https://database.cosmin.nl/.

46. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159-70.

47. DeVellis RF. Scale development: Theory and applications. 4th Ed. ed. Bickman L, Rog DJ, editors. Thousand Oaks, California: SAGE Publications, Inc.; 2017. 262 p.

48. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. Health Econ. 1999;8(1):41-51.

49. Babbie ER. The Practice of Social Research. Boston, MA, UNITED STATES: Cengage Learning US; 2015.

50. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155-63.

51. Streiner DL, Norman GR. Health Measurement Scales: A practical guide to their development and use. Oxford Scholarship Online: Oxford Scholarship Online; 2008.

52. American Psychological Association. Reliability Washington, DC: APA; 2020 [cited July 2021]. Available from: https://dictionary.apa.org/reliability.

53. Yong AG, Pearce S. A Beginner's guide to factor analysis: Focusing on Exploratory Factor Analysis. Tutor Quant Methods Psychol. 2013;9(2):79-94.

54. Gartner FR, Bomhof-Roordink H, Smith IP, Scholl I, Stiggelbout AM, Pieterse AH. The quality of instruments to assess the process of shared decision making: A systematic review. PLoS One. 2018;13(2):e0191747.

55. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.

56. Maternity Choices Australia. About Brisbane, Australia2021 [cited 2021 October]. Available from: https://www.maternitychoices.org/about.

SUPPLEMENTARY FILE 1 – MEDLINE (OVID) ELECTRONIC DATABASE SEARCH STRATEGIES

Maternity PROMs (07-10-2021)

n	**Search terms**	**Results**
1	((Obstetrics/) OR (Prenatal Care/) OR (Maternal Health Services/) OR (Pregnancy/) OR (Delivery, Obstetric/) OR (Postnatal Care/) OR (Pregnant women/) OR (Parturition/) OR (Labor, Obstetric/) OR (Delivery, obstetric/) OR (Obstetric?.ab,ti.) OR (Matern$.ab,ti.) (Prenatal.ab,ti.) OR (Antepartum.ab,ti.) OR (Perinatal.ab,ti.) OR (Intrapartum.ab,ti.) OR (Postnatal.ab,ti.) OR (Postpartum.ab,ti.) OR (Pregnan$.ab,ti.) OR (Childbirth?.ab,ti.) OR (Birth$.ab,ti.))	1,356,332
2	((Patient Reported Outcome Measures/) OR ((patient reported outcome?).ab,ti.) OR ((patient-reported outcome?).ab,ti.) OR (("patient-centred outcome?" or "patient centred outcome?").ti,ab.) OR (("patient-centered outcome?" or "patient centered outcome?").ti,ab.))	22,352
3	(((instrumentation or methods).fs.) OR (("Validation Studies" or "Comparative Study").pt.) OR (Psychometrics/) OR (psychometr$.ti,ab.) OR ((clinimetr$ or clinometr$).tw.) OR (Outcome Assessment, Health Care/) OR (outcome measure$.tw.) OR (Observer Variation/) OR (observer variation.ti,ab.) OR (Health Status Indicators/) OR (Reproducibility of Results/) OR (reproducib$.ti,ab.) OR (Discriminant Analysis/) OR (outcome assessment$.ti,ab.) OR ((reliab$ or unreliab$ or valid$ or coefficient$ or homogeneity or homogeneous or internal consistency).ti,ab.) OR ((cronbach$ and alpha$).ti,ab.) OR ((item and (correlation$ or selection$ or reduction$)).ti,ab.) OR ((agreement or precision or imprecision or "precise values" or test$retest).ti,ab.) OR (test retest.ti,ab.) OR ((reliab$ and (test or retest)).ti,ab.) OR ((stability or inter$rater or intra$rater or inter$tester or intra$tester or inter$observer or intra$observer or inter$technician or intra$technician or inter$examiner or intra$examiner or inter$assay or intra$assay or inter$individual or intra$individual or inter$participant or intra$participant or kappa$ or repeatab$).ti,ab.) OR (((replicab$ or repeated) and (measure$ or finding$ or result$ or test$)).ti,ab.) OR ((generaliza$ or generalisa$ or concordance).ti,ab.) OR (intraclass correlation$.ti,ab.) OR ((discriminative or "known group" or "factor analys$" or dimension$ or subscale$).ti,ab.) OR (multitrait scaling analys$.ti,ab.) OR (("item discriminant" or "interscale correlation$" or error$ or "individual variability").ti,ab.) OR ((variability and (analysis or values)).ti,ab.) OR ((uncertainty and (measurement or measuring)).ti,ab.) OR (("standard error of measurement" or sensitiv$ or responsive$).ti,ab.) OR (((minimal$ or clinical$) and (important or significant or detectable) and (change or difference)).ti,ab.) OR ((small$ and (real or detectable) and (change or difference)).ti,ab.) OR (("meaningful change" or "ceiling effect$" or "floor effect$" or "Item response model$" or IRT or Rasch or "Differential item functioning" or DIF or "computer adaptive testing" or "item bank" or "cross$cultural equivalence").ti,ab.))	8,796,309
4	(("Surveys and Questionnaires"/) OR (questionnaire?.ab,ti.) OR (tool.ab,ti.) OR (scale?.ab,ti.) OR (measure$.ab,ti.) OR (instrument?.ab,ti.))	4,331,661
5	(2 AND 3 AND 4)	14,185
6	(1 AND 5) Limited to: English language; Publication year 2010- current	**111**

? = Wildcard that stands for 0 or 1 replacement character within or at the end of a word; $ = Wildcard that searches for variations of a word of unlimited character length; / = Medical Subject Heading (MeSH) term; ti = Title search; ab = Abstract search

Maternity PREMs (07-10-2021)

n	Search terms	Results
1	((Obstetrics/) OR (Prenatal Care/) OR (Maternal Health Services/) OR (Pregnancy/) OR (Delivery, Obstetric/) OR (Postnatal Care/) OR (Pregnant women/) OR (Parturition/) OR (Labor, Obstetric/) OR (Delivery, obstetric/) OR (Obstetric?.ab,ti.) OR (Matern$.ab,ti.) (Prenatal.ab,ti.) OR (Antepartum.ab,ti.) OR (Perinatal.ab,ti.) OR (Intrapartum.ab,ti.) OR (Postnatal.ab,ti.) OR (Postpartum.ab,ti.) OR (Pregnan$.ab,ti.) OR (Childbirth?.ab,ti.) OR (Birth$.ab,ti.))	1,356,332
2	((“patient reported experience?”.ab,ti.) OR (“patient-reported experience?”.ab,ti.) OR ((wom?n? experience?).ab,ti.) OR ((woman$ experience?).ab,ti.))	9,041
3	(((instrumentation or methods).fs.) OR (("Validation Studies" or "Comparative Study").pt.) OR (Psychometrics/) OR (psychometr$.ti,ab.) OR ((clinimetr$ or clinometr$).tw.) OR (Outcome Assessment, Health Care/) OR (outcome measure$.tw.) OR (Observer Variation/) OR (observer variation.ti,ab.) OR (Health Status Indicators/) OR (Reproducibility of Results/) OR (reproducib$.ti,ab.) OR (Discriminant Analysis/) OR (outcome assessment$.ti,ab.) OR ((reliab$ or unreliab$ or valid$ or coefficient$ or homogeneity or homogeneous or internal consistency).ti,ab.) OR ((cronbach$ and alpha$).ti,ab.) OR ((item and (correlation$ or selection$ or reduction$)).ti,ab.) OR ((agreement or precision or imprecision or "precise values" or test$retest).ti,ab.) OR (test retest.ti,ab.) OR ((reliab$ and (test or retest)).ti,ab.) OR ((stability or inter$rater or intra$rater or inter$tester or intra$tester or inter$observer or intra$observer or inter$technician or intra$technician or inter$examiner or intra$examiner or inter$assay or intra$assay or inter$individual or intra$individual or inter$participant or intra$participant or kappa$ or repeatab$).ti,ab.) OR (((replicab$ or repeated) and (measure$ or finding$ or result$ or test$)).ti,ab.) OR ((generaliza$ or generalisa$ or concordance).ti,ab.) OR (intraclass correlation$.ti,ab.) OR ((discriminative or "known group" or "factor analys$" or dimension$ or subscale$).ti,ab.) OR (multitrait scaling analys$.ti,ab.) OR (("item discriminant" or "interscale correlation$" or error$ or "individual variability").ti,ab.) OR ((variability and (analysis or values)).ti,ab.) OR ((uncertainty and (measurement or measuring)).ti,ab.) OR (("standard error of measurement" or sensitiv$ or responsive$).ti,ab.) OR (((minimal$ or clinical$) and (important or significant or detectable) and (change or difference)).ti,ab.) OR ((small$ and (real or detectable) and (change or difference)).ti,ab.) OR (("meaningful change" or "ceiling effect$" or "floor effect$" or "Item response model$" or IRT or Rasch or "Differential item functioning" or DIF or "computer adaptive testing" or "item bank" or "cross$cultural equivalence").ti,ab.))	8,796,309
4	(("Surveys and Questionnaires"/) OR (questionnaire?.ab,ti.) OR (tool.ab,ti.) OR (scale?.ab,ti.) OR (measure$.ab,ti.) OR (instrument?.ab,ti.))	4,331,661
5	(2 AND 3 AND 4)	1,754
6	(1 AND 5) Limited to: English language; Publication year 2010- current	**496**

? = Wildcard that stands for 0 or 1 replacement character within or at the end of a word; $ = Wildcard that searches for variations of a word of unlimited character length; / = Medical Subject Heading (MeSH) term; ti = Title search; ab = Abstract search

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Checklist item	Page No
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	Identify the report as a protocol of a systematic review	1
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	n/a
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	2
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	1
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	22
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	n/a
Support:			
Sources	5a	Indicate sources of financial or other support for the review	22
Sponsor	5b	Provide name for the review funder and/or sponsor	22
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	22
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	4-5
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	5-6
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	6-7
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	8
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	Appendix 1

Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	8
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	8-9
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	9, 13-14
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	9-14
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	9-14
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	9-13
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	9-14
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)	15
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	n/a
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	n/a
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	n/a
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	n/a

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.

BMJ Open

Evaluating the development, woman-centricity and psychometric properties of maternity patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs): A systematic review protocol

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-058952.R1
Article Type:	Protocol
Date Submitted by the Author:	07-Dec-2021
Complete List of Authors:	Bull, Claudia; Monash University, Monash Centre for Health Research and Implementation; Monash University, School of Public Health and Preventive Medicine Teede, Helena; Monash University, Monash Centre for Health Research and Implementation; Monash University, Monash Partners Academic Health Science Centre Carrandi, Lane; Monash University, Monash Centre for Health Research and Implementation; Monash University, School of Public Health and Preventive Medicine Rigney, Azure; Maternity Choices Australia Cusack, Sally; Maternity Choices Australia Callander, Emily; Monash University, Monash Centre for Health Research and Implementation; Monash University, School of Public Health and Preventive Medicine
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health policy, Research methods, Obstetrics and gynaecology, Patient-centred medicine
Keywords:	HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, OBSTETRICS

Evaluating the development, woman-centricity and psychometric properties of maternity patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs): A systematic review protocol

Claudia Bull[1,2], Helena Teede[1,3], Lane Carrandi[1,2], Azure Rigney[4], Sally Cusack[4], Emily J Callander[1,2]

[1]Monash Centre for Health Research and Implementation, Monash University, 43-51 Kanooka Grove, Clayton, VIC, 3168, AUSTRALIA

[2]School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, VIC, 3004, AUSTRALIA

[3]Monash Partners Academic Health Science Centre, 43-51 Kanooka Grove, Clayton, VIC, 3168, AUSTRALIA

[4]Maternity Choices Australia, Springwood, Queensland, 4127, AUSTRALIA

Corresponding author:

Miss Claudia Bull

School of Public Health and Preventive Medicine, Monash University

Level 4, 553 St Kilda Road, Melbourne, VIC, 3004, AUSTRALIA

E: claudia.bull@monash.edu

Ph: +61 (03) 9903 0723

Word count: 3,121

ABSTRACT

Introduction: Woman-centred care is the right of every woman receiving maternity care, irrespective of where care is being received and who is providing care. This protocol describes a planned systematic review that will identify, describe, and critically appraise the psychometric properties of maternity patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs). The woman-centricity of PROM and PREM development and content validation (i.e., the extent to which women were involved in these processes) will also be assessed. This information will be used to develop a maternity PROMs and PREMs database to support service and system performance measurement, and value-based maternity care initiatives.

Methods and analysis: This study will be guided by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Guideline for Systematic Reviews of Outcome Measurement Instruments. Studies identified via MEDLINE, CINAHL Plus, PsycINFO, and EMBASE describing the development, content validation, and/or psychometric evaluation of PROMs and PREMs specifically designed for maternity populations throughout pregnancy, childbirth and postnatal periods will be considered if published from 2010 onwards, in English, and available in full-text. The COSMIN risk of bias checklist will be used to evaluate the quality of studies reporting on the development, content validation, and/or psychometric evaluation of PROMs and PREMs. COSMIN criteria for good content validity will be used to assess the woman-centricity of PROM and PREM development and content validation studies. COSMIN standards of good psychometric properties will be used to evaluate the validity and reliability of the identified instruments.

Ethics and dissemination: Ethical permission for this research is not required. The findings of this research will be submitted for publication in an international, peer-reviewed journal. The proposed maternity PROMs and PREMs database will be freely accessible online, and developed with consumer input to ensure its usefulness to a range of maternity care stakeholders.

PROSPERO registration: CRD42021288854

Keywords: Patient-reported outcome measure (PROM); Patient-reported experience measure (PREM); Woman-centred care; Survey development; Psychometric evaluation; Validity; Reliability

STRENGTHS AND LIMITATIONS OF THIS STUDY

- Employing the COSMIN guidance at all stages provides a transparent, uniform and robust approach to the conduct of this systematic review.
- Compiling evidence on the woman-centricity (i.e., the involvement of women) in instrument development and content validation is yet to be evidenced in the peer-reviewed literature, and aims to support performance measurement and value-based assessment that is meaningful to women.
- Developing a publicly available database of maternity PROMs and PREMs aims to promote best practice instrument selection and implementation to support the measurement of services and systems, and contribute to operationalising value-based health care.
- A potential limitation of this review is using COSMIN guidance (developed for PROMs) to evaluate the development, content validation and psychometric evaluation of PREMs.
- Additionally, the review will only include PROMs and PREMs published after 2010, and studies published in English.

INTRODUCTION

The concept of woman-centred care (WCC) is underpinned by the principles of choice, control, continuity of carer, and a woman's right to self-determination.(1-3) WCC is typically associated with midwifery practice,(4) but this misrepresents the reality that receiving WCC is the right of every woman, irrespective of where or by whom she receives care. Coupled with a "risk avoidant" obstetric culture and increasing rates of intervention at birth (particularly in high-income countries),(3, 5, 6) women's values and preferences for aspects of care beyond a successful live birth (e.g., desire for a natural birth) are often a secondary consideration. This has subsequently challenged the implementation of value-based maternity care, where consumer perspectives are at the centre of outcome measurement.

Value-based healthcare (VBHC) is the purported goal of every health system. At its core, VBHC aims to improve patient health outcomes relative to the cost of achieving those improvements.(7) However, VBHC frameworks that exist on this principle alone have been called into question as they oversimplify the complex construct of 'value';(8) particularly what value means to patients in different circumstances.(9) Indeed, in the context of maternity care, women value a diverse array of factors including care continuity, equitability, promoting normal reproductive processes, choosing where they give birth, being treated respectfully, emotional support, and transparent communication.(10-13) Consequently, value-based maternity care represents far more than a successful live birth.

One means of capturing the experiences and outcomes of maternity care that women value, is using patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs). Despite capturing different elements of healthcare encounters, both types of instrument are designed to measure and evaluate service and system performance from the consumer's perspective.(14) By responding to the outcomes and experiences reported by consumers, health services and systems are better able to support VBHC. However, this is only achieved if the content of PROMs and PREMs aligns with what is viewed as important and relevant to care consumers (i.e., women). Thus, woman-centric instrument development and content validation – that is, the involvement of women in

defining what is relevant, comprehensive and comprehendible instrument content – is crucial to supporting meaningful, value-based measurement in maternity care.(15)

Specifically, PROMs measure an individual's health and wellbeing.(16, 17) They can capture a wide-range of outcomes, largely related to physical, social, and/or psychological functioning.(16, 18, 19) Recent reviews of condition-specific PROMs used during pregnancy and childbirth have revealed instruments capturing recovery after childbirth,(20) outpatient postpartum recovery,(21) sleep in postpartum women,(22) postpartum pain,(23) and functional recovery following caesarean section.(24) However, PROMs capturing outcomes relevant to all women across the pregnancy, childbirth and postpartum continuum are missing.(25)

PREMs differ in that they are designed to capture an individual's experience of receiving care; namely, what happened during a care encounter and their perception of how it happened.(26) There are no reviews of maternity PREMs, however, a recent concept analysis identified several constructs commonly captured in relation to women's experiences of maternity care. These include organisational aspects of care such as access and referral to maternity services, continuity of care, privacy and care costs; and interpersonal aspects of care such as information sharing, informed choice, emotional support, being treated with respect and dignity, and having confidence in the knowledge and ability of maternity care providers.(27)

We intend to develop a database hosting a repository of PROMS and PREMs to support the use of these instruments in health services and systems performance measurement and evaluation as a part of achieving value-based maternity care. Specifically, we aim to identify and appraise PROMs and PREMs that capture outcomes and experiences (respectively) relevant to all women across the pregnancy, childbirth and postpartum continuum. This protocol describes the systematic process that will be undertaken to firstly, identify and describe maternity PROMs and PREMs published in the peer-reviewed literature, and secondly, critically appraise and summarise the psychometric properties of the identified instruments. Particular emphasis will be placed on the woman-centeredness of PROM and PREM development. The database will subsequently summarise this information in a user-friendly format suitable for a range of maternity care stakeholders.

METHODS AND ANALYSIS

This study will be guided by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) *Guideline for Systematic Reviews of Outcome Measurement Instruments*.(28) COSMIN stipulates a 10-step process for performing a systematic review of PROMs (which will be extended to PREMs for the purposes of this research). Steps 1-4 pertain to conducting the literature search; steps 5-7 pertain to evaluating an instruments' psychometric properties; step 8 pertains to evaluating the interpretability and feasibility of implementing instruments; and steps 9-10 pertain to writing the review discussion. This protocol will detail the processes we intend to undertake for steps 1-8.

Step 1: Formulate the aim of the review

The aim of this review is two-fold. First, to identify and describe maternity PROMs and PREMs relevant to all women across the pregnancy, childbirth and postpartum continuum, published in the peer-reviewed literature. Second, to critically appraise and summarise the psychometric properties of the identified instruments, with particular emphasis on assessing the woman-centricity of instrument development and content validation.

Step 2: Formulate the eligibility criteria

The eligibility criteria in **Table 1** will be applied:

Table 1: Systematic review eligibility criteria for studies reporting on maternity PROMs and PREMs

PROM studies will be included if:	PREM studies will be included if:
• Published from 2010 onwards, representing contemporary instruments (however, if articles refer to earlier papers describing developmental and psychometric evaluation evidence pre-dating 2010, we will include these to provide a holistic representation of instrument quality) • Published in English • Available in full-text • Described the development, content validation and/ or psychometric evaluation of PROMs relevant to all women across the pregnancy, childbirth and postpartum continuum	• Published from 2010 onwards, representing contemporary instruments (however, if articles refer to earlier papers describing developmental and psychometric evaluation evidence pre-dating 2010, we will include these to provide a holistic representation of instrument quality) • Published in English • Available in full-text • Described the development, content validation and/ or psychometric evaluation of PREMs relevant to all women receiving maternity care
PROM studies will be excluded if:	**PREM studies will be excluded if:**
• Published before 2010 (except as specified above) • Published in languages other than English • Not available in full-text • Presented literature reviews, meta-reviews, protocols, theses or quality improvement activities • The included instruments were not clearly PROMs (e.g., Body Experience during Pregnancy Scale [BEPS](29)) • Included PROMs were used as outcome measures (e.g., in an RCT), but did not contribute to their development, content validation, and/ or psychometric evaluation • Described proxy-reported PROMs (i.e., not self-reported by women)	• Published before 2010 (except as specified above) • Published in languages other than English • Not available in full-text • Presented literature reviews, meta-reviews, protocols, theses or quality improvement activities • The included instruments that were not clearly PREMs • Included PREMs were used as outcome measures (e.g., in a cross-sectional study), but did not contribute to their development, content validation, and/ or psychometric evaluation • Described proxy-reported PREMs (i.e., not self-reported by women)

- Described the development, content validation and/ or psychometric evaluation of:
  - Generic PROMs (e.g., Patient-Reported Measurement Information System [PROMIS](30))
  - PROMs originally developed in contexts other than maternity (e.g., Postnatal Demoralisation Scale [Postnatal DS](31))
  - Quality of life instruments/ utility measures (e.g., Labor and Delivery Index [LADY-X](32); Mother Generated Index [MGI](33))
  - PROMs for specific maternal sub-populations (e.g., Pelvic Girdle Questionnaire [PGQ](34))
  - Screening tools (e.g., Edinburgh Postnatal Depression Scale(35))
  - Core outcome sets(36, 37)

- Described the development, content validation and/ or psychometric evaluation of:
  - Satisfaction or expectation measures (e.g., the Birth Satisfaction Scale [BSS](38-40))[a]
  - PREMs originally developed in a context other than maternity (e.g., inpatient, outpatient settings)
  - PREMs for specific maternal sub-populations (e.g., women receiving abortion care(41))

[a]When there was ambiguity between satisfaction measures and PREMs, we considered: (i) the instruments' response scale (noting that agreement-based scales are more common in satisfaction measures whereas frequency-based scales are more common in PREMs), (ii) whether questions were expectation-based (aligning with satisfaction), and (iii) whether the original intent behind instrument development was to measure satisfaction or experiences;(42) PROM = patient-reported outcome measure; PREM = patient-reported experience measure; RCT = randomised controlled trial

We specifically delineate PROMs from quality of life/ utility measures. Quality of life/ utility measures (e.g., EQ-5D, Health Utility Index and SF-6D) are preference-based instruments despite often being referred to interchangeably as PROMs.(43-46) While PROMs and quality of life/ utility measures may capture similar constructs,(47) they differ in how they are used and scored. PROMs were originally developed with the intent to assess health *outcomes* with reference to receipt of healthcare, and are scored on an item-by-item or domain/dimension basis. Conversely, quality of life/ utility measures were originally developed for the purposes of quantifying a person's health state (without specific reference to receiving healthcare) and their present level of quality of life. Thus, quality of life/utility measures are used to determine quality adjusted life years (QALYs), where an individuals' quality of life as it relates to their health state is scored as one of a finite number of health states relative to a utility index.(47-50) As such, they will not be included in this review.

We also specifically delineate PREMs from patient satisfaction measures, despite being referred to synonymously throughout the literature.(42) PREMs ask individuals to *report* on their experiences of care, where satisfaction measures ask individuals to *evaluate* their experiences. While report style questions aim to be objective,(16) evaluative questions are more likely to reflect an individuals' expectations, attitudes and desire to appear socially desirable, and are thus influenced by attributes peripheral to their care experience.(51-53) Additionally, where PREMs typically use frequency-based response scales (e.g., on a scale of never to always),(54-56) patient satisfaction measures tend to use agreement-based response scales (e.g., on a scale of strongly disagree to strongly agree),(57-59) which are more prone to acquiescence bias and straightlining.(60, 61) Thus, given the differences between these instruments, patient satisfaction measures will not be included in this review.

Step 3: Perform a literature search

We will search the following electronic databases: MEDLINE (via Ovid), CINAHL Plus (via EBSCOhost), PsycINFO (via Ovid), and EMBASE (via Elsevier). Our search terms will include the following concepts: (i) maternity care and maternal health services; (ii) PROMs; (iii) PREMs; and (iv) measurement properties. We will employ the search terms developed by COSMIN relevant to studies on measurement properties.(62) These are available for each of MEDLINE, CINAHL, PsycINFO and

EMABSE. An example of our proposed MEDLINE search strategy is available in **Supplementary file 1**. Searches will be limited to only studies published in English and available in full-text.

Step 4: Select abstracts and full-text articles

After being exported from electronic databases, all search results will be imported into Covidence.(63) Two reviewers will independently review all titles and abstracts to determine which articles warrant full-text retrieval and review. Full text review will also be undertaken by two independent reviewers. Discrepancies at all stages will be addressed through reviewer consultation and consensus, and if needed, engagement of a third reviewer. The reference lists of all included papers will be hand-searched for other potentially relevant studies.(28)

Steps 5-7: Evaluating the measurement properties of the included PROMs and PREMs

Data extraction – Characteristics of the included PROMs and PREMs

The following data will be extracted from included studies: (i) PROM/ PREM name; (ii) construct(s)/ domain(s) captured; (iii) target population and setting; (iv) mode of administration (e.g., online, postal), and administration time during perinatal care (e.g. antenatal, postnatal); (v) recall period; (vi) number of items; (vii) response options; and (viii) original language. This information will be used to describe the included PROMs and PREMs. Information will be extracted per study and grouped where there have been multiple studies conducted for one instrument. One reviewer will extract all data.

Evaluating the methodological quality of studies

Methodological quality will be evaluated in relation to maternity PROM and PREM development, content validation, and psychometric evaluation using the COSMIN Risk of Bias checklist.(64) This checklist details specific study design elements that are important when assessing the measurement properties of an instrument. Only study design elements relevant to the measurement properties presented in **Table 2** and reported in studies will be assessed for risk of bias. Criteria for study design elements are rated using a scale of 'very good', 'adequate', 'doubtful', inadequate', or 'n/a'. The

lowest rating for any criteria will be used to describe the quality of the study underpinning that specific measurement property (i.e., worst score counts).(28) If multiple studies have been conducted to evidence a specific measurement property (e.g., three studies report on an instruments' internal consistency) and have provided variable results, the overall quality of the measurement property will be labelled 'unclear'.

One reviewer will first consult the *COSMIN database of systematic reviews of outcome measurement instruments*(65) to determine whether other researchers have already evaluated the risk of bias of the included studies. If available, existing ratings will be used (as is recommended by COSMIN(28)). If not, or if additional evidence supporting an instrument has been published, one reviewer will determine the measurement property(ies) that need to be assessed per study, and two reviewers will independently complete the Risk of Bias checklist for each individual study. We will use the Risk of Bias Microsoft Excel template developed by COSMIN to document each rater's scores.

Table 2: Elements of the COSMIN Risk of Bias checklist(64) for assessing study design relative to PROM and PREM development, content validation and psychometric evaluation studies

Measurement property	Number of criteria
Content validity	
PROM/ PREM development	35
Content validity	31
Internal structure	
Structural validity	4
Internal consistency	5
Cross-cultural validity/ measurement in variance	4
Remaining measurement properties	
Reliability	8
Measurement error	6
Criterion validity	3
Hypotheses testing for construct validity	7
Responsiveness	13

PROM = Patient-reported outcome measure; PREM = Patient-reported experience measure

Evaluating the content validity (woman-centricity) of PROM and PREM development

Content validity has been described as the most important measurement property of PROMs (and arguably, PREMs).(66) It represents the degree to which the content of an instrument is an adequate reflection of the phenomena being measured.(67) PROM and PREM items need to demonstrate appropriate relevance, comprehensiveness and comprehensibility to qualify as content valid.(66) This assessment should be made by 'experts' of the target phenomena. In the context of maternity care, the women receiving and experiencing care are the experts. COSMIN also provide criteria to support studies that have asked health professionals about the relevance and comprehensiveness of items.(66) Instruments that fail to demonstrate appropriate involvement of women in their development and content validation will be labelled as demonstrating 'inadequate' content validity.

COSMIN has developed a set of instructions specifically for evaluating the content validity of PROMs which will be used in this study (for both PROMs and PREMs). The first two steps involve evaluating the quality of studies reporting on instrument development and content validation; this forms part of the COSMIN Risk of Bias checklist described above. The third step involves rating each development and content validation study against nine criteria for good content validity (**Table 3**).(66) For each of relevance, comprehensiveness and comprehensibility, if ≥85% of an instruments' items fulfil the criteria, the study is deemed to have sufficient (+) evidence; if <85% of items fulfil the criteria, the study is deemed to have insufficient (—) evidence; and if there is inadequate information available or the study quality was inadequate (as identified through risk of bias assessment), the study is deemed to have indeterminate (?) evidence.(66) From this, we will assign an overall content validation score (+, —, ?) which will represent the woman-centricity of PROM and PREM development. Two reviewers will undertake the content validation assessment.

Table 3: Relevance, comprehensiveness and comprehensibility criteria for evaluating the content validity of maternity care instruments(66)

Criteria
Relevance
1. Are the included items relevant to maternity care?
2. Are the included items relevant to women?

3. Are the response options appropriate?
4. Is the recall period appropriate?
Comprehensiveness
5. Are all key concepts included?
Comprehensibility
6. Are the instrument instructions understood by women as intended?
7. Are items and response options understood by women as intended?
8. Are items appropriately worded?
9. Do the response options match the question?

Evaluating the sufficiency of measurement properties

Instruments will next be evaluated according to how well the reported measurement properties (e.g., structural validity) comply with standards of good psychometric properties (**Table 4**).(28) This will indicate whether a PROM or PREM can be considered valid and reliable. Validity is the extent to which an instrument measures what is purports to measure.(68, 69) Reliability is the extent to which participant responses to an instrument can be replicated in unchanging circumstances (consistency).(70) Reliability is also the extent to which an instrument is devoid of measurement error.(71, 72)

Using the COSMIN updated criteria for good measurement properties, psychometric properties will be rated as + (provides sufficient evidence), ─ (provides insufficient evidence), and ? (provides indeterminate evidence) (**Table 4)**. Red text denotes criteria added based on prominence in the literature relative to instrument development and psychometric evaluation. COSMIN's criteria of 'Hypothesis testing for construct validity' has been excluded from Table 3 as the context of maternity care in this study is too broad for the review team to appropriately generate hypotheses suitable for all potential instruments. If a PROM or PREM has several studies reporting on its' psychometric properties, each study will be evaluated individually (according to the reported psychometric properties), and an overall conclusion regarding the quality of the instrument will be provided for each psychometric quality. Any psychometric properties not assessed will be labelled as having 'no evidence'. Two reviewers will undertake the good psychometric properties assessment.

Table 4: COSMIN updated criteria for good measurement properties(28)

Measurement property	Rating	Criteria
Structural validity	+	**Classical test theory (CTT)** • Confirmatory factor analysis (CFA): CFI or TLI (or comparable measure) >0.95, **OR** RMSEA <0.06, **OR** SMSR <0.08; **AND/OR** • Exploratory factor analysis (EFA) or Principal Components Analysis (PCA): KMO ≥0.70, **AND** Significant Bartlett's Test of Sphericity (p<0.05), **AND** dimensional (total) variance explained ≥50% or dimensional (total) variance explained <50% but justified by the authors(73, 74) **Item Response Theory (IRT)/ Rasch:** • No violation of unidimensionality: CFI or TLI (or comparable measure) >0.95, **OR** RMSEA <0.06, **OR** SMSR <0.08; **AND** • No violation of local independence: residual correlations among the items after controlling for the dominant factor <0.20 OR Q_3 fit statistics <0.37; **AND** • No violation of monotonicity: adequate looking graphs, OR item scalability >0.30; **AND** • Adequate model fit – IRT: χ^2 >0.01; Rasch: infit and outfit mean squares between ≥0.50 and ≤1.50, OR z-standardised values between >-2 and <2
	?	**CTT** • Not all information for + reported **IRT/ Rasch** • Model fit not reported
	–	**CTT** • Criteria for + not achieved **IRT/ Rasch** • Criteria for + not achieved
Internal consistency	+	Evidence of sufficient structural validity achieved (+ or ? for 'Structural validity'); **AND** Cronbach's alpha(s) ≥0.70 for each unidimensional scale or subscale
	?	Evidence of sufficient structural validity not achieved
	–	Evidence of sufficient structural validity achieved (+ or ? for 'Structural validity'); **AND** Cronbach's alpha(s) <0.70 for each unidimensional scale or subscale
Reliability	+	ICC or weighted Kappa ≥0.70
	?	ICC or weighted Kappa not reported
	–	ICC or weighted Kappa <0.70
Measurement error	+	SDC or LoA < MIC
	?	MIC not defined
	–	SDC or LoA > MIC
Cross-cultural validity/ measurement invariance	+	No importance differences found between group factors (such as age, gender, language) in multiple group factor analysis; **OR** No important DIF for group factors (McFadden's R^2 <0.02)
	?	No multiple group factor analysis performed; **OR** No DIF analysis performed
	–	Important differences between group factor analysis identified; **OR** Important differences in DIF analysis identified
Criterion validity	+	Correlation with gold standard instrument ≥0.70*; **OR** AUC ≥0.70
	?	Not all information for + reported
	–	Correlation with gold standard instrument <0.70*; **OR** AUC <0.70
Responsiveness	+	AUC ≥0.70
	?	AUC not reported
	–	AUC <0.70

*Correlation with a gold standard will only occur if a short-form instrument is being compared against its long-form counterpart; CTT = Classical Test Theory; CFA = Confirmatory Factor Analysis; CFI = Comparative Fit Index;/ TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SMSR = Standardised Root Mean Residuals; EFA = Exploratory Factor Analysis; CA = Principal Components Analysis; KMO = Kaiser-Meyer-Olkin; IRT = Item Response Theory; ICC = Intraclass Correlation Coefficient; SDC = Smallest Detectable Change; LoA = Limits of Agreement; MIC = Minimally Important Change; DIF = Differential Item Functioning; Area Under the Curve

Summarise and grade the quality of evidence

By summarising and grading the evidence available for an individual instrument, we can provide an overall conclusion as to the quality of that instrument. Thus, this will involve combining the results of each instruments' risk of bias, content validity, and psychometric property assessments into a single metric of 'high', 'moderate', 'low', or 'very low' evidence using the Grading of Recommendations Assessment, Development, and Evaluation (GRAE) approach.(28) If the results across multiple studies pertaining to a single instrument are consistent, then results will be quantitatively pooled and a GRADE score will be reported. If results are inconsistent, they will not be pooled, no GRADE score will be reported, and areas of inconsistency will be discussed (e.g., if an instrument demonstrates differing levels of quality depending on the country in which it is used).

Step 8: Describe the interpretability and feasibility of instrument implementation

Interpretability is the extent to which meaning can be derived from participant responses to an instrument or changes in responses.(75) This may include distinct patterns of responses amongst subgroups of the population, trends in responses over time, and floor or ceiling effects. For the purposes of this review, we will extract and describe the following features of PROM and PREM interpretability: (i) distribution of responses in the study population and relevant subgroups; (ii) proportion of missing data for items; (iii) methods of handling missing data; (iv) evidence of floor and ceiling effects; and (v) minimally important changes (MIC) or minimally important differences (MID) in responses. Interpretability, whilst not considered a measurement property in and of itself, is important for understanding the real-world application and biases associated with implementing PROMs and PREMs.

Feasibility refers to the ease and convenience with which a PROM or PREM can be implemented and administered in a real-world context.(28) For the purposes of this review, we will extract and describe the following features of PROM and PREM feasibility: (i) available modes of administration; (ii) length of the instrument; (iii) estimated completion time; (iv) level of readability; (v) ease of response calculation; (vi) copyright; (vii) cost of using an instrument; (viii) equipment required for instrument administration; (ix) availability of instrument for application in different settings and languages; and (x) approvals required before instrument use. For the development of a maternity PROMs and PREMs database, this information will be critical for informing the real-world implementation of maternity PROMs and PREMs across health services and systems.

Patient and public involvement statement

The research team comprises members of Maternity Choices Australia, a national consumer advocacy organisation committed to the advancement of best-practice maternity care.(76) These women are consumer representatives and have been involved in the conceptualisation of the research and protocol development, recognising the importance of operationalising woman-centred care, and ensuring that maternity services are consumer informed. Importantly, they will aid the development of the Maternity PROMs and PREMs database, supporting its usability by a range of maternity care stakeholders. They will also help disseminate the Maternity PROMs and PREMs database through formal and informal engagement with key collaborative parties.

ETHICS AND DISSEMINATION

Ethical permission for this research is not required as the review will only use information from previously published research. The findings of this research will be submitted for publication in an international, peer-reviewed journal. Abstracts will also be submitted for national and international conference presentations.

Maternity PROMs and PREMs database

We intend for the maternity PROMs and PREMs database to be freely accessible online, and useful to all individuals involved in maternity health services and systems performance measurement, and

value-based maternity care. The design of the database will be consumer informed to ensure that it is easy to understand, and provides information relevant to a range of maternity care stakeholders. The psychometric results (structural validity, internal consistency, reliability, measurement error, cross-cultural validity/ measurement invariance, criterion validity and responsiveness) for each instrument will be summarised according to whether criteria were met (+), indeterminate (?), or not met (—) when all evidence for a specific instrument in considered collectively. The woman-centricity of instrument development will be similarly summarised according to the COSMIN criteria for good content validity. In addition, the database will summarise descriptive information for each instrument (e.g., number of items, domains captured, country of development); summarise information regarding each instruments' feasibility of use (e.g., copyright and reuse considerations, available modes of administration, costs etc.); and provide links to the studies describing instruments. For PROMs or PREMs not freely available, we will also provide the appropriate contact information for the instruments' original author or licensing agent.

We anticipate that the database will be updated annually. A member of the research team will re-run the search strategies (updating search terms as needed) and undertake the processes described in this protocol. This will support the identification of new instruments or additional evidence of PROM and PREM psychometric evaluation over time, ensuring that the database is up-to-date and aligns with advancements in PROM and PREM methodologies and results.

REFERENCES

1. Leap N. Woman-centred or women-centred care: Does it matter? British Journal of Midwifery. 2009;17(1):12-6.

2. Snowden A, Martin C, Jomeen J, Hollins Martin C. Concurrent analysis of choice and control in childbirth. BMC Pregnancy Childbirth. 2011;11:40.

3. Brady S, Lee N, Gibbons K, Bogossian F. Woman-centred care: An integrative review of the empirical literature. Int J Nurs Stud. 2019;94:107-19.

4. Fontein-Kuipers Y, de Groot R, van Staa A. Woman-centered care 2.0: Bringing the concept into focus. Eur J Midwifery. 2018;2(May).

5. Akileswaran CP, Hutchison MS. Making Room at the Table for Obstetrics, Midwifery, and a Culture of Normalcy Within Maternity Care. Obstet Gynecol. 2016;128(1):176-80.

6. Maffi I, Gouilhers S. Conceiving of risk in childbirth: Obstetric discourses, medical management and cultural expectations in Switzerland and Jordan. Health Risk Soc. 2019;21(3-4):185-206.

7. Porter ME, Teisberg EO. Redefining health care: creating value-based competition on results. Brighton, MA: Harvard business press; 2006.

8. Perfetto EM. ISPOR's Initiative on US Value Assessment Frameworks: A Missed Opportunity for ISPOR and Patients. Value Health. 2018;21(2):169-70.

9. Perfetto EM, Oehrlein EM, Boutin M, Reid S, Gascho E. Value to Whom? The Patient Voice in the Value Discussion. Value Health. 2017;20(2):286-91.

10. De Jonge A, Downe S, Page L, Devane D, Lindgren H, Klinkert J, et al. Value based maternal and newborn care requires alignment of adequate resources with high value activities. BMC Pregnancy Childbirth. 2019;19(1):428.

11. Fawsitt CG, Bourke J, Greene RA, McElroy B, Krucien N, Murphy R, et al. What do women want? Valuing women's preferences and estimating demand for alternative models of maternity care using a discrete choice experiment. Health Policy. 2017;121(11):1154-60.

12. Fawsitt CG, Bourke J, Lutomski JE, Meaney S, McElroy B, Murphy R, et al. What women want: Exploring pregnant women's preferences for alternative models of maternity care. Health Policy. 2017;121(1):66-74.

13. Renfrew MJ, McFadden A, Bastos MH, Campbell J, Channon AA, Cheung NF, et al. Midwifery and quality care: Findings from a new evidence-informed framework for maternal and newborn care. Lancet. 2014;384(9948):1129-45.

14. Organisation for Economic Co-operation and Development. Measuring what matters: the patient-reported indicator surveys. Paris: OECD; 2019.

15. Terwee CB, Prinsen CAC, Chiarotto A, de Vet HCW, Bouter LM, Alonso J, et al. COSMIN methodology for assessing the content validity and PROMs: User manual. Amsterdam: COSMIN; 2018.

16. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. Bja Educ. 2017;17(4):137-44.

17. Black N, Burke L, Forrest CB, Sieberer UH, Ahmed S, Valderas JM, et al. Patient-reported outcomes: pathways to better health, better services, and better societies. Qual Life Res. 2016;25(5):1103-12.

18. Williams K, Sansoni J, Morris D, Grootemaat P, Thompson C. Patient-reported outcome measures: Literature review. Sydney: ACSQHC; 2016 November 2016.

19. Centre for Health Services and Policy Research. What are PROMs? Vancouver: The University of British Columbia; [cited 2021 June]. Available from: https://patientreportedoutcomes.ca/what-are-pros/proms/.

20. Sultan P, Sadana N, Sharawi N, Blake L, El-Boghdadly K, Falvo A, et al. Evaluation of Domains of Patient-Reported Outcome Measures for Recovery After Childbirth: A Scoping and Systematic Review. JAMA Netw Open. 2020;3(5):e205540.

21. Sultan P, Sharawi N, Blake L, Ando K, Sultan E, Aghaeepour N, et al. Use of Patient-Reported Outcome Measures to Assess Outpatient Postpartum Recovery: A Systematic Review. JAMA Netw Open. 2021;4(5):e2111600.

22. Sultan P, Ando K, Sultan E, Hawkins J, Blake L, Barwick F, et al. A systematic review of patient-reported outcome measures used to assess sleep in postpartum women using Consensus Based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines. Sleep. 2021;44(10).

23. Sultan P, Ando K, Sultan E, Hawkins JE, Chitneni A, Sharawi N, et al. A systematic review of patient-reported outcome measures to assess postpartum pain using Consensus Based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines. Br J Anaesth. 2021;127(2):264-74.

24. Sharawi N, Klima L, Shah R, Blake L, Carvalho B, Sultan P. Evaluation of patient-reported outcome measures of functional recovery following caesarean section: a systematic review using the consensus-based standards for the selection of health measurement instruments (COSMIN) checklist. Anaesthesia. 2019;74(11):1439-55.

25. Dickinson F, McCauley M, Smith H, van den Broek N. Patient reported outcome measures for use in pregnancy and childbirth: a systematic review. Bmc Pregnancy Childb. 2019;19.

26. Bull C, Byrnes J, Hettiarachchi R, Downes M. A systematic review of the validity and reliability of patient-reported experience measures. Health Serv Res. 2019;54(5):1023-35.

27. Beecher C, Devane D, White M, Greene R, Dowling M. Women's experiences of their maternity care: A principle- based concept analysis. Women Birth. 2020;33(5):419-25.

28. Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, de Vet HCW, et al. COSMIN methodology for systematic reivews of patient-reported outcome measures (PROMs). Amsterdam: COSMIN; 2018.

29. Talmon A, Ginzburg K. "Who does this body belong to?" The development and psychometric evaluation of the Body Experience during Pregnancy Scale. Body Image. 2018;26:19-28.

30. Lundsberg LS, Schwarz EB, Vilardo NA, Yonkers KA, Gariepy AM. Clinical Validation of PROMIS Global Short Form in Pregnancy. Appl Res Qual Life. 2018;13(1):89-103.

31. Bobevski I, Rowe H, Clarke DM, McKenzie DP, Fisher J. Postnatal demoralisation among women admitted to a hospital mother-baby unit: validation of a psychometric measure. Archives of Women's Mental Health. 2015;18(6):817-27.

32. Gartner FR, de Miranda E, Rijnders ME, Freeman LM, Middeldorp JM, Bloemenkamp KW, et al. Good reliability and validity for a new utility instrument measuring the birth experience, the Labor and Delivery Index. J Clin Epidemiol. 2015;68(10):1184-94.

33. Symon A, Downe S, Finlayson KW, Knapp R, Diggle P, team St. The feasibility and acceptability of using the Mother-Generated Index (MGI) as a Patient Reported Outcome Measure in a randomised controlled trial of maternity care. Bmc Med Res Methodol. 2015;15:100.

34. Stuge B, Jenssen HK, Grotle M. The Pelvic Girdle Questionnaire: Responsiveness and Minimal Important Change in Women With Pregnancy-Related Pelvic Girdle Pain, Low Back Pain, or Both. Phys Ther. 2017;97(11):1103-13.

35. Cox JL, Holden JM, Sagovsky R. Detection of Postnatal Depression: Development of the 10-item Edinburgh Postnatal Depression Scale. British Journal of Psychiatry. 1987;150(6):782-6.

36. Schaap T, Bloemenkamp K, Deneux-Tharaux C, Knight M, Langhoff-Roos J, Sullivan E, et al. Defining definitions: a Delphi study to develop a core outcome set for conditions of severe maternal morbidity. BJOG: An International Journal of Obstetrics & Gynaecology. 2019;126(3):394-401.

37. Briscoe KE, Haas DM. Developing a Core Outcome Set for Cesarean Delivery Maternal Infectious Morbidity Outcomes. Am J Perinatol. 2020;37(4):436-52.

38. Martin CH, Fleming V. The birth satisfaction scale. Int J Health Care Qual Assur. 2011;24(2):124-35.

39. Vardavaki Z, Hollins Martin CJ, Martin CR. Construct and content validity of the Greek version of the Birth Satisfaction Scale (G-BSS). Journal of Reproductive and Infant Psychology. 2015;33(5):488-503.

40. Göncü Serhatlıoğlu S, Karahan N, Hollins Martin CJ, Martin CR. Construct and content validity of the Turkish Birth Satisfaction Scale - Revised (T-BSS-R). J Reprod Infant Psychol. 2018;36(3):235-45.

41. Sudhinaraset M, Landrian A, Afulani PA, Phillips B, Diamond-Smith N, Cotter S. Development and validation of a person-centered abortion scale: the experiences of care in private facilities in Kenya. BMC Womens Health. 2020;20(1):208.

42. Bull C. Patient satisfaction and patient experience are not interchangeable concepts. Int J Qual Health Care. 2021;33(1).

43. Jonsson A, Orwelius L, Dahlstrom U, Kristenson M. Evaluation of the usefulness of EQ-5D as a patient-reported outcome measure using the Paretian classification of health change among patients with chronic heart failure. J Patient Rep Outcomes. 2020;4(1):50.

44. Benson T, Williams DH, Potts HW. Performance of EQ-5D, howRu and Oxford hip & knee scores in assessing the outcome of hip and knee replacements. BMC Health Serv Res. 2016;16(1):512.

45. NHS Digital. Patient Reported Outcome Measures (PROMs) in England - 2011-2012, Special topic, EQ-5D™ Index analysis UK: NHS; 2014 [updated 12 October 2018; cited 2021 June]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms/patient-reported-outcome-measures-proms-in-england-2011-2012-special-topic-eq-5d-index-analysis.

46. Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Serv Insights. 2013;6:61-8.

47. Feeny DH, Eckstrom E, Whitlock EP, Perdue LA. Patient-Reported Outcomes, Health-Related Quality of Life, and Function: An Overview of Measurement Properties. A Primer for Systematic Reviewers on the Measurement of Functional Status and Health-Related Quality of Life in Older Adults. AHRQ Methods for Effective Health Care. Rockville (MD): AHRQ; 2013.

48. Whitehurst DG, Noonan VK, Dvorak MF, Bryan S. A review of preference-based health-related quality of life questionnaires in spinal cord injury research. Spinal Cord. 2012;50(9):646-54.

49. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. Br Med Bull. 2010;96:5-21.

50. Jain SS, DeFroda SF, Paxton ES, Green A. Patient-Reported Outcome Measures and Health-Related Quality-of-Life Scores of Patients Undergoing Anatomic Total Shoulder Arthroplasty. J Bone Joint Surg Am. 2020;102(13):1186-93.

51. Williams B, Coyle J, Healy D. The meaning of patient satisfaction: an explanation of high reported levels. Soc Sci Med. 1998;47(9):1351-9.

52. Noyes RW, Levy MI, Chase CL, Udry JR. Expectation fulfillment as a measure of patient satisfaction. American Journal of Obstetrics and Gynecology. 1974;118(6):809-14.

53. Williams B. Patient satisfaction: a valid concept? Soc Sci Med. 1994;38(4):509-16.

54. Centers for Medicare & Medicaid Services. HCAHPS Hospital Consumer Assessment of Healthcare Providers and Systems Baltimore, MD: CMS; 2021 [updated 2 February 2021; cited 2021 June]. Available from: https://hcahpsonline.org/en/survey-instruments/.

55. U.S. Centers for Medicare & Medicaid Services. Emergency Department CAHPS (ED CAHPS) Baltimore, MD: CMS.gov; 2020 [updated 11 May 2020; cited 2021 June]. Available from: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/CAHPS/ED.

56. Care Quality Commission. Adult Inpatient Survey 2020 UK: NHS; 2020 [cited 2021 June]. Available from: https://nhssurveys.org/surveys/survey/02-adults-inpatients/.

57. Grogan S, Conner M, Willits D, Norman P. Development of a questionnaire to measure patients' satisfaction with general practitioners' services. Br J Gen Pract. 1995;45(399):525-9.

58. Kinnersley P, Stott N, Peters T, Harvey I, Hackett P. A comparison of methods for measuring patient satisfaction with consultations in primary care. Fam Pract. 1996;13(1):41-51.

59. Hojat M, Louis DZ, Maxwell K, Markham FW, Wender RC, Gonnella JS. A brief instrument to measure patients' overall satisfaction with primary care physicians. Fam Med. 2011;43(6):412-7.

60. Lavrakas PJ. Acquiescence Response Bias. In: Lavrakas PJ, editor. Encyclopedia of Survey Research Methods. Thousand Oaks, CA: Sage Publications; 2008.

61. Dykema J, Schaeffe NC, Garbarski D, Assad N, Blixt S. Towards a reconsideration of the use of agree-disagree questions in measuring subjective evaluations. Res Social Adm Pharm. 2021.

62. Terwee CB, Jansma EP, Riphagen, II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009;18(8):1115-23.

63. Covidence.org. Covidence Melbourne: Covidence.org; 2021 [cited 2021 September]. Available from: https://www.covidence.org/.

64. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018;27(5):1171-9.

65. COSMIN. COSMIN database of systematic reviews of outcome measurement instruments Amsterdam: University Library Vrje Universiteit Amsterdam; 2019 [updated Septmeber 2019; cited 2021 September]. Available from: https://database.cosmin.nl/.

66. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159-70.

67. DeVellis RF. Scale development: Theory and applications. 4th Ed. ed. Bickman L, Rog DJ, editors. Thousand Oaks, California: SAGE Publications, Inc.; 2017. 262 p.

68. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. Health Econ. 1999;8(1):41-51.

69. Babbie ER. The Practice of Social Research. Boston, MA, UNITED STATES: Cengage Learning US; 2015.

70. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155-63.

71. Streiner DL, Norman GR. Health Measurement Scales: A practical guide to their development and use. Oxford Scholarship Online: Oxford Scholarship Online; 2008.

72. American Psychological Association. Reliability Washington, DC: APA; 2020 [cited July 2021]. Available from: https://dictionary.apa.org/reliability.

73. Yong AG, Pearce S. A Beginner's guide to factor analysis: Focusing on Exploratory Factor Analysis. Tutor Quant Methods Psychol. 2013;9(2):79-94.

74. Gartner FR, Bomhof-Roordink H, Smith IP, Scholl I, Stiggelbout AM, Pieterse AH. The quality of instruments to assess the process of shared decision making: A systematic review. PLoS One. 2018;13(2):e0191747.

75. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.

76. Maternity Choices Australia. About Brisbane, Australia2021 [cited 2021 October]. Available from: https://www.maternitychoices.org/about.

Funding statement: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests: Nil.

Author contributions: CB contributed to the conception of the research and wrote the protocol manuscript. HT and LC provided important intellectual content in revisions of the protocol manuscript. AR and SC provided important consumer representation and content relative to woman-centred care. EC contributed to the conception of the research and provided important intellectual content in revisions of the protocol manuscript. All authors approved of the version submitted for publication and agree to be accountable for all aspects of the research.

Data sharing: Not applicable.

SUPPLEMENTARY FILE 1 – MEDLINE (OVID) ELECTRONIC DATABASE SEARCH STRATEGIES

Maternity PROMs (07-10-2021)

n	**Search terms**	**Results**
1	((Obstetrics/) OR (Prenatal Care/) OR (Maternal Health Services/) OR (Pregnancy/) OR (Delivery, Obstetric/) OR (Postnatal Care/) OR (Pregnant women/) OR (Parturition/) OR (Labor, Obstetric/) OR (Delivery, obstetric/) OR (Obstetric?.ab,ti.) OR (Matern$.ab,ti.) (Prenatal.ab,ti.) OR (Antepartum.ab,ti.) OR (Perinatal.ab,ti.) OR (Intrapartum.ab,ti.) OR (Postnatal.ab,ti.) OR (Postpartum.ab,ti.) OR (Pregnan$.ab,ti.) OR (Childbirth?.ab,ti.) OR (Birth$.ab,ti.))	1,356,332
2	((Patient Reported Outcome Measures/) OR ((patient reported outcome?).ab,ti.) OR ((patient-reported outcome?).ab,ti.) OR (("patient-centred outcome?" or "patient centred outcome?").ti,ab.) OR (("patient-centered outcome?" or "patient centered outcome?").ti,ab.))	22,352
3	(((instrumentation or methods).fs.) OR (("Validation Studies" or "Comparative Study").pt.) OR (Psychometrics/) OR (psychometr$.ti,ab.) OR ((clinimetr$ or clinometr$).tw.) OR (Outcome Assessment, Health Care/) OR (outcome measure$.tw.) OR (Observer Variation/) OR (observer variation.ti,ab.) OR (Health Status Indicators/) OR (Reproducibility of Results/) OR (reproducib$.ti,ab.) OR (Discriminant Analysis/) OR (outcome assessment$.ti,ab.) OR ((reliab$ or unreliab$ or valid$ or coefficient$ or homogeneity or homogeneous or internal consistency).ti,ab.) OR ((cronbach$ and alpha$).ti,ab.) OR ((item and (correlation$ or selection$ or reduction$)).ti,ab.) OR ((agreement or precision or imprecision or "precise values" or test$retest).ti,ab.) OR (test retest.ti,ab.) OR ((reliab$ and (test or retest)).ti,ab.) OR ((stability or inter$rater or intra$rater or inter$tester or intra$tester or inter$observer or intra$observer or inter$technician or intra$technician or inter$examiner or intra$examiner or inter$assay or intra$assay or inter$individual or intra$individual or inter$participant or intra$participant or kappa$ or repeatab$).ti,ab.) OR (((replicab$ or repeated) and (measure$ or finding$ or result$ or test$)).ti,ab.) OR ((generaliza$ or generalisa$ or concordance).ti,ab.) OR (intraclass correlation$.ti,ab.) OR ((discriminative or "known group" or "factor analys$" or dimension$ or subscale$).ti,ab.) OR (multitrait scaling analys$.ti,ab.) OR (("item discriminant" or "interscale correlation$" or error$ or "individual variability").ti,ab.) OR ((variability and (analysis or values)).ti,ab.) OR ((uncertainty and (measurement or measuring)).ti,ab.) OR (("standard error of measurement" or sensitiv$ or responsive$).ti,ab.) OR (((minimal$ or clinical$) and (important or significant or detectable) and (change or difference)).ti,ab.) OR ((small$ and (real or detectable) and (change or difference)).ti,ab.) OR (("meaningful change" or "ceiling effect$" or "floor effect$" or "Item response model$" or IRT or Rasch or "Differential item functioning" or DIF or "computer adaptive testing" or "item bank" or "cross$cultural equivalence").ti,ab.))	8,796,309
4	(("Surveys and Questionnaires"/) OR (questionnaire?.ab,ti.) OR (tool.ab,ti.) OR (scale?.ab,ti.) OR (measure$.ab,ti.) OR (instrument?.ab,ti.))	4,331,661
5	(2 AND 3 AND 4)	14,185
6	(1 AND 5) Limited to: English language; Publication year 2010- current	**111**

? = Wildcard that stands for 0 or 1 replacement character within or at the end of a word; $ = Wildcard that searches for variations of a word of unlimited character length; / = Medical Subject Heading (MeSH) term; ti = Title search; ab = Abstract search

Maternity PREMs (07-10-2021)

n	Search terms	Results
1	((Obstetrics/) OR (Prenatal Care/) OR (Maternal Health Services/) OR (Pregnancy/) OR (Delivery, Obstetric/) OR (Postnatal Care/) OR (Pregnant women/) OR (Parturition/) OR (Labor, Obstetric/) OR (Delivery, obstetric/) OR (Obstetric?.ab,ti.) OR (Matern$.ab,ti.) (Prenatal.ab,ti.) OR (Antepartum.ab,ti.) OR (Perinatal.ab,ti.) OR (Intrapartum.ab,ti.) OR (Postnatal.ab,ti.) OR (Postpartum.ab,ti.) OR (Pregnan$.ab,ti.) OR (Childbirth?.ab,ti.) OR (Birth$.ab,ti.))	1,356,332
2	((“patient reported experience?”.ab,ti.) OR (“patient-reported experience?”.ab,ti.) OR ((wom?n? experience?).ab,ti.) OR ((woman$ experience?).ab,ti.))	9,041
3	(((instrumentation or methods).fs.) OR (("Validation Studies" or "Comparative Study").pt.) OR (Psychometrics/) OR (psychometr$.ti,ab.) OR ((clinimetr$ or clinometr$).tw.) OR (Outcome Assessment, Health Care/) OR (outcome measure$.tw.) OR (Observer Variation/) OR (observer variation.ti,ab.) OR (Health Status Indicators/) OR (Reproducibility of Results/) OR (reproducib$.ti,ab.) OR (Discriminant Analysis/) OR (outcome assessment$.ti,ab.) OR ((reliab$ or unreliab$ or valid$ or coefficient$ or homogeneity or homogeneous or internal consistency).ti,ab.) OR ((cronbach$ and alpha$).ti,ab.) OR ((item and (correlation$ or selection$ or reduction$)).ti,ab.) OR ((agreement or precision or imprecision or "precise values" or test$retest).ti,ab.) OR (test retest.ti,ab.) OR ((reliab$ and (test or retest)).ti,ab.) OR ((stability or inter$rater or intra$rater or inter$tester or intra$tester or inter$observer or intra$observer or inter$technician or intra$technician or inter$examiner or intra$examiner or inter$assay or intra$assay or inter$individual or intra$individual or inter$participant or intra$participant or kappa$ or repeatab$).ti,ab.) OR (((replicab$ or repeated) and (measure$ or finding$ or result$ or test$)).ti,ab.) OR ((generaliza$ or generalisa$ or concordance).ti,ab.) OR (intraclass correlation$.ti,ab.) OR ((discriminative or "known group" or "factor analys$" or dimension$ or subscale$).ti,ab.) OR (multitrait scaling analys$.ti,ab.) OR (("item discriminant" or "interscale correlation$" or error$ or "individual variability").ti,ab.) OR ((variability and (analysis or values)).ti,ab.) OR ((uncertainty and (measurement or measuring)).ti,ab.) OR (("standard error of measurement" or sensitiv$ or responsive$).ti,ab.) OR (((minimal$ or clinical$) and (important or significant or detectable) and (change or difference)).ti,ab.) OR ((small$ and (real or detectable) and (change or difference)).ti,ab.) OR (("meaningful change" or "ceiling effect$" or "floor effect$" or "Item response model$" or IRT or Rasch or "Differential item functioning" or DIF or "computer adaptive testing" or "item bank" or "cross$cultural equivalence").ti,ab.))	8,796,309
4	(("Surveys and Questionnaires"/) OR (questionnaire?.ab,ti.) OR (tool.ab,ti.) OR (scale?.ab,ti.) OR (measure$.ab,ti.) OR (instrument?.ab,ti.))	4,331,661
5	(2 AND 3 AND 4)	1,754
6	(1 AND 5) Limited to: English language; Publication year 2010- current	**496**

? = Wildcard that stands for 0 or 1 replacement character within or at the end of a word; $ = Wildcard that searches for variations of a word of unlimited character length; / = Medical Subject Heading (MeSH) term; ti = Title search; ab = Abstract search

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Checklist item	Page No
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	Identify the report as a protocol of a systematic review	1
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	n/a
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	2
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	1
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	26
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	n/a
Support:			
Sources	5a	Indicate sources of financial or other support for the review	26
Sponsor	5b	Provide name for the review funder and/or sponsor	26
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	26
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	4-5
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	5-6
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	7-8
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	9-10

Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	Supplementary file 1
Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	10
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	10
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	10-15
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	10-15
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	10-15
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	10-15
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	10-15
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)	15
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	n/a
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	n/a
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	n/a
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	n/a

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.

BMJ Open

Evaluating the development, woman-centricity and psychometric properties of maternity patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs): A systematic review protocol

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-058952.R2
Article Type:	Protocol
Date Submitted by the Author:	28-Dec-2021
Complete List of Authors:	Bull, Claudia; Monash University, Monash Centre for Health Research and Implementation; Monash University, School of Public Health and Preventive Medicine Teede, Helena; Monash University, Monash Centre for Health Research and Implementation; Monash University, Monash Partners Academic Health Science Centre Carrandi, Lane; Monash University, Monash Centre for Health Research and Implementation; Monash University, School of Public Health and Preventive Medicine Rigney, Azure; Maternity Choices Australia Cusack, Sally; Maternity Choices Australia Callander, Emily; Monash University, Monash Centre for Health Research and Implementation; Monash University, School of Public Health and Preventive Medicine
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health policy, Research methods, Obstetrics and gynaecology, Patient-centred medicine
Keywords:	HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, OBSTETRICS

Evaluating the development, woman-centricity and psychometric properties of maternity patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs): A systematic review protocol

Claudia Bull[1,2], Helena Teede[1,3], Lane Carrandi[1,2], Azure Rigney[4], Sally Cusack[4], Emily J Callander[1,2]

[1]Monash Centre for Health Research and Implementation, Monash University, 43-51 Kanooka Grove, Clayton, VIC, 3168, AUSTRALIA

[2]School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, VIC, 3004, AUSTRALIA

[3]Monash Partners Academic Health Science Centre, 43-51 Kanooka Grove, Clayton, VIC, 3168, AUSTRALIA

[4]Maternity Choices Australia, Springwood, Queensland, 4127, AUSTRALIA

Corresponding author:

Miss Claudia Bull

School of Public Health and Preventive Medicine, Monash University

Level 4, 553 St Kilda Road, Melbourne, VIC, 3004, AUSTRALIA

E: claudia.bull@monash.edu

Ph: +61 (03) 9903 0723

Word count: 3,265

ABSTRACT

Introduction: Woman-centred care is the right of every woman receiving maternity care, irrespective of where care is being received and who is providing care. This protocol describes a planned systematic review that will identify, describe, and critically appraise the psychometric properties of maternity patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs). The woman-centricity of PROM and PREM development and content validation (i.e., the extent to which women were involved in these processes) will also be assessed. This information will be used to develop a maternity PROMs and PREMs database to support service and system performance measurement, and value-based maternity care initiatives.

Methods and analysis: This study will be guided by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Guideline for Systematic Reviews of Outcome Measurement Instruments. Studies identified via MEDLINE, CINAHL Plus, PsycINFO, and EMBASE describing the development, content validation, and/or psychometric evaluation of PROMs and PREMs specifically designed for maternity populations throughout pregnancy, childbirth and postnatal periods will be considered if published from 2010 onwards, in English, and available in full-text. The COSMIN risk of bias checklist will be used to evaluate the quality of studies reporting on the development, content validation, and/or psychometric evaluation of PROMs and PREMs. COSMIN criteria for good content validity will be used to assess the woman-centricity of PROM and PREM development and content validation studies. COSMIN standards of good psychometric properties will be used to evaluate the validity and reliability of the identified instruments.

Ethics and dissemination: Ethical permission for this research is not required. The findings of this research will be submitted for publication in an international, peer-reviewed journal. Abstracts for national and international conference presentations will also be submitted. The proposed maternity PROMs and PREMs database will be freely accessible online, and developed with consumer input to ensure its usefulness to a range of maternity care stakeholders.

PROSPERO registration: CRD42021288854

Keywords: Patient-reported outcome measure (PROM); Patient-reported experience measure (PREM); Woman-centred care; Survey development; Psychometric evaluation; Validity; Reliability

STRENGTHS AND LIMITATIONS OF THIS STUDY

- Employing the COSMIN guidance at all stages provides a transparent, uniform and robust approach to the conduct of this systematic review.
- Compiling evidence on the woman-centricity (i.e., the involvement of women) in instrument development and content validation is yet to be evidenced in the peer-reviewed literature, and aims to support performance measurement and value-based assessment that is meaningful to women.
- Developing a publicly available database of maternity PROMs and PREMs aims to promote best practice instrument selection and implementation to support the measurement of services and systems, and contribute to operationalising value-based health care.
- A potential limitation of this review is using COSMIN guidance (developed for PROMs) to evaluate the development, content validation and psychometric evaluation of PREMs.
- Additionally, the review will only include PROMs and PREMs published after 2010, and studies published in English.

INTRODUCTION

The concept of woman-centred care (WCC) is underpinned by the principles of choice, control, continuity of carer, and a woman's right to self-determination.(1-3) WCC is typically associated with midwifery practice,(4) but this misrepresents the reality that receiving WCC is the right of every woman, irrespective of where or by whom she receives care. Coupled with a "risk avoidant" obstetric culture and increasing rates of intervention at birth (particularly in high-income countries),(3, 5, 6) women's values and preferences for aspects of care beyond a successful live birth (e.g., desire for a natural birth) are often a secondary consideration. This has subsequently challenged the implementation of value-based maternity care, where consumer perspectives are at the centre of outcome measurement.

Value-based healthcare (VBHC) is the purported goal of every health system. At its core, VBHC aims to improve patient health outcomes relative to the cost of achieving those improvements.(7) However, VBHC frameworks that exist on this principle alone have been called into question as they oversimplify the complex construct of 'value';(8) particularly what value means to patients in different circumstances.(9) Indeed, in the context of maternity care, women value a diverse array of factors including care continuity, equitability, promoting normal reproductive processes, choosing where they give birth, being treated respectfully, emotional support, and transparent communication.(10-13) Consequently, value-based maternity care represents far more than a successful live birth.

One means of capturing the experiences and outcomes of maternity care that women value, is using patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs). Despite capturing different elements of healthcare encounters, both types of instrument are designed to measure and evaluate service and system performance from the consumer's perspective.(14) By responding to the outcomes and experiences reported by consumers, health services and systems are better able to support VBHC. However, this is only achieved if the content of PROMs and PREMs aligns with what is viewed as important and relevant to care consumers (i.e., women). Thus, woman-centric instrument development and content validation – that is, the involvement of women in

defining what is relevant, comprehensive and comprehendible instrument content – is crucial to supporting meaningful, value-based measurement in maternity care.(15)

Specifically, PROMs measure an individual's health and wellbeing.(16, 17) They can capture a wide-range of outcomes, largely related to physical, social, and/or psychological functioning.(16, 18, 19) Recent reviews of condition-specific PROMs used during pregnancy and childbirth have revealed instruments capturing recovery after childbirth,(20) outpatient postpartum recovery,(21) sleep in postpartum women,(22) postpartum pain,(23) and functional recovery following caesarean section.(24) However, PROMs capturing outcomes relevant to all women across the pregnancy, childbirth and postpartum continuum are missing.(25)

PREMs differ in that they are designed to capture an individual's experience of receiving care; namely, what happened during a care encounter and their perception of how it happened.(26) There are no reviews of maternity PREMs, however, a recent concept analysis identified several constructs commonly captured in relation to women's experiences of maternity care. These include organisational aspects of care such as access and referral to maternity services, continuity of care, privacy and care costs; and interpersonal aspects of care such as information sharing, informed choice, emotional support, being treated with respect and dignity, and having confidence in the knowledge and ability of maternity care providers.(27)

We intend to develop a database hosting a repository of PROMS and PREMs to support the use of these instruments in health services and systems performance measurement and evaluation as a part of achieving value-based maternity care. Specifically, we aim to identify and appraise PROMs and PREMs that capture outcomes and experiences (respectively) relevant to maternity care that is accessed by all women across the pregnancy, childbirth and postpartum continuum. This protocol describes the systematic process that will be undertaken to firstly, identify and describe maternity PROMs and PREMs published in the peer-reviewed literature, and secondly, critically appraise and summarise the psychometric properties of the identified instruments. Particular emphasis will be placed on the woman-centeredness of PROM and PREM development. The database will

subsequently summarise this information in a user-friendly format suitable for a range of maternity care stakeholders.

METHODS AND ANALYSIS

This study will be guided by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) *Guideline for Systematic Reviews of Outcome Measurement Instruments*.(28) COSMIN stipulates a 10-step process for performing a systematic review of PROMs (which will be extended to PREMs for the purposes of this research). Steps 1-4 pertain to conducting the literature search; steps 5-7 pertain to evaluating an instruments' psychometric properties; step 8 pertains to evaluating the interpretability and feasibility of implementing instruments; and steps 9-10 pertain to writing the review discussion. This protocol will detail the processes we intend to undertake for steps 1-8.

Step 1: Formulate the aim of the review

The aim of this review is two-fold. First, to identify and describe maternity PROMs and PREMs relevant to all women across the pregnancy, childbirth and postpartum continuum, published in the peer-reviewed literature. Second, to critically appraise and summarise the psychometric properties of the identified instruments, with particular emphasis on assessing the woman-centricity of instrument development and content validation.

Step 2: Formulate the eligibility criteria

The following eligibility criteria will be applied:

Table 1: Systematic review eligibility criteria for studies reporting on maternity PROMs and PREMs

PROM studies will be included if:	PREM studies will be included if:
• Published from 2010 onwards, representing contemporary instruments (however, if articles refer to earlier papers describing developmental and psychometric evaluation evidence pre-dating 2010, we will include these to provide a holistic representation of instrument quality) • Published in English • Available in full-text • Described the development, content validation and/ or psychometric evaluation of PROMs relevant to all women across the pregnancy, childbirth and postpartum continuum	• Published from 2010 onwards, representing contemporary instruments (however, if articles refer to earlier papers describing developmental and psychometric evaluation evidence pre-dating 2010, we will include these to provide a holistic representation of instrument quality) • Published in English • Available in full-text • Described the development, content validation and/ or psychometric evaluation of PREMs relevant to all women receiving maternity care
PROM studies will be excluded if:	**PREM studies will be excluded if:**
• Published before 2010 (except as specified above) • Published in languages other than English • Not available in full-text • Presented literature reviews, meta-reviews, protocols, theses or quality improvement activities • The included instruments were not clearly PROMs (e.g., Body Experience during Pregnancy Scale [BEPS](29)) • Included PROMs were used as outcome measures (e.g., in an RCT), but did not contribute to their development, content validation, and/ or psychometric evaluation • Described proxy-reported PROMs (i.e., not self-reported by women)	• Published before 2010 (except as specified above) • Published in languages other than English • Not available in full-text • Presented literature reviews, meta-reviews, protocols, theses or quality improvement activities • The included instruments that were not clearly PREMs • Included PREMs were used as outcome measures (e.g., in a cross-sectional study), but did not contribute to their development, content validation, and/ or psychometric evaluation • Described proxy-reported PREMs (i.e., not self-reported by women)

- Described the development, content validation and/ or psychometric evaluation of:
  - Generic PROMs (e.g., Patient-Reported Measurement Information System [PROMIS](30))
  - PROMs originally developed in contexts other than maternity (e.g., Postnatal Demoralisation Scale [Postnatal DS](31))
  - Quality of life instruments/ utility measures (e.g., Labor and Delivery Index [LADY-X](32); Mother Generated Index [MGI](33))
  - PROMs for specific maternal sub-populations (e.g., Pelvic Girdle Questionnaire [PGQ](34))
  - Screening tools (e.g., Edinburgh Postnatal Depression Scale(35))
  - Core outcome sets(36, 37)

- Described the development, content validation and/ or psychometric evaluation of:
  - Satisfaction or expectation measures (e.g., the Birth Satisfaction Scale [BSS](38-40))[a]
  - PREMs originally developed in a context other than maternity (e.g., inpatient, outpatient settings)
  - PREMs for specific maternal sub-populations (e.g., women receiving abortion care(41))

[a]When there was ambiguity between satisfaction measures and PREMs, we considered: (i) the instruments' response scale (noting that agreement-based scales are more common in satisfaction measures whereas frequency-based scales are more common in PREMs), (ii) whether questions were expectation-based (aligning with satisfaction), and (iii) whether the original intent behind instrument development was to measure satisfaction or experiences;(42) PROM = patient-reported outcome measure; PREM = patient-reported experience measure; RCT = randomised controlled trial

We specifically delineate PROMs from quality of life/ utility measures. Quality of life/ utility measures (e.g., EQ-5D, Health Utility Index and SF-6D) are preference-based instruments despite often being referred to interchangeably as PROMs.(43-46) While PROMs and quality of life/ utility measures may capture similar constructs,(47) they differ in how they are used and scored. PROMs were originally developed with the intent to assess health *outcomes* with reference to receipt of health care, and are scored on an item-by-item or domain/dimension basis. Conversely, quality of life/ utility measures were originally developed for the purposes of quantifying a person's health state (without any reference to having received health care or not) and their present level of quality of life. Furthermore, quality of life/utility measures can then be used for determining quality adjusted life years (QALYs) where an individuals' quality of life as it relates to their health state is scored as one of a finite number of health states relative to a utility index.(47-50) As such, they will not be included in this review.

We also specifically delineate PREMs from patient satisfaction measures, despite being referred to synonymously throughout the literature.(42) PREMs ask individuals to *report* on their experiences of care, where satisfaction measures ask individuals to *evaluate* their experiences. While report style questions aim to be objective,(16) evaluative questions are more likely to reflect an individuals' expectations, attitudes and desire to appear socially desirable, and are thus influenced by attributes peripheral to their care experience.(51-53) Additionally, where PREMs typically use frequency-based response scales (e.g., on a scale of never to always),(54-56) patient satisfaction measures tend to use agreement-based response scales (e.g., on a scale of strongly disagree to strongly agree),(57-59) which are more prone to acquiescence bias and straightlining.(60, 61) Thus, given the differences between these instruments, patient satisfaction measures will not be included in this review.

Step 3: Perform a literature search

We will search the following electronic databases: MEDLINE (via Ovid), CINAHL Plus (via EBSCOhost), PsycINFO (via Ovid), and EMBASE (via Elsevier). Our search terms will include the following concepts: (i) maternity care and maternal health services; (ii) PROMs; (iii) PREMs; and (iv) measurement properties. We will employ the search terms developed by COSMIN relevant to studies

on measurement properties.(62) These are available for each of MEDLINE, CINAHL, PsycINFO and EMABSE. An example of our proposed MEDLINE search strategy is available in **Supplementary file 1**. Searches will be limited to only studies published in English and available in full-text.

Step 4: Select abstracts and full-text articles

After being exported from electronic databases, all search results will be imported into Covidence.(63) Two reviewers will independently review all titles and abstracts to determine which articles warrant full-text retrieval and review. Full text review will also be undertaken by two independent reviewers. Discrepancies at all stages will be addressed through reviewer consultation and consensus, and if needed, engagement of a third reviewer. The reference lists of all included papers will be hand-searched for other potentially relevant studies.(28)

Steps 5-7: Evaluating the measurement properties of the included PROMs and PREMs

Data extraction – Characteristics of the included PROMs and PREMs

The following data will be extracted from included studies: (i) PROM/ PREM name; (ii) construct(s)/ domain(s) captured; (iii) target population and setting; (iv) mode of administration (e.g., online, postal), and administration time during perinatal care (e.g. antenatal, postnatal); (v) recall period; (vi) number of items; (vii) response options; and (viii) original language. This information will be used to describe the included PROMs and PREMs. Information will be extracted per study and grouped where there have been multiple studies conducted for one instrument. One reviewer will extract all data.

Evaluating the methodological quality of studies

Methodological quality will be evaluated in relation to maternity PROM and PREM development, content validation, and psychometric evaluation using the COSMIN Risk of Bias checklist.(28) This checklist details specific study design elements that are important when assessing the measurement properties of an instrument. Only study design elements relevant to the measurement properties presented in **Table 1** and reported in studies will be assessed for risk of bias. Criteria for study design

elements are rated using a scale of 'very good', 'adequate', 'doubtful', inadequate', or 'n/a'. The lowest rating for any criteria will be used to describe the quality of the study underpinning that specific measurement property (i.e., worst score counts).(28) If multiple studies have been conducted to evidence a specific measurement property (e.g., three studies report on an instruments' internal consistency) and have provided variable results, the overall quality of the measurement property will be labelled 'unclear'.

One reviewer will first consult the *COSMIN database of systematic reviews of outcome measurement instruments*(64) to determine whether other researchers have already evaluated the risk of bias of the included studies. If available, existing ratings will be used (as is recommended by COSMIN(28)). If not, or if additional evidence supporting an instrument has been published, one reviewer will determine the measurement property(ies) that need to be assessed per study, and two reviewers will independently complete the Risk of Bias checklist for each individual study. We will use the Risk of Bias Microsoft Excel template developed by COSMIN to document each rater's scores.

Table 1: Elements of the COSMIN Risk of Bias checklist(28) for assessing study design relative to PROM and PREM development, content validation and psychometric evaluation studies

Measurement property	Number of criteria
Content validity	
PROM/ PREM development	35
Content validity	31
Internal structure	
Structural validity	4
Internal consistency	5
Cross-cultural validity/ measurement in variance	4
Remaining measurement properties	
Reliability	8
Measurement error	6
Criterion validity	3
Hypotheses testing for construct validity	7
Responsiveness	13

PROM = Patient-reported outcome measure; PREM = Patient-reported experience measure

Evaluating the content validity (woman-centricity) of PROM and PREM development

Content validity has been described as the most important measurement property of PROMs (and arguably, PREMs).(65) It represents the degree to which the content of an instrument is an adequate reflection of the phenomena being measured.(66) PROM and PREM items need to demonstrate appropriate relevance, comprehensiveness and comprehensibility to qualify as content valid.(65) This assessment should be made by 'experts' of the target phenomena. In the context of maternity care, the women receiving and experiencing care are the experts. COSMIN also provide criteria to support studies that have asked health professionals about the relevance and comprehensiveness of items.(65) Instruments that fail to demonstrate appropriate involvement of women in their development and content validation will be labelled as demonstrating 'inadequate' content validity.

COSMIN has developed a set of instructions specifically for evaluating the content validity of PROMs which will be used in this study (for both PROMs and PREMs). The first two steps involve evaluating the quality of studies reporting on instrument development and content validation; this forms part of the COSMIN Risk of Bias checklist described above. The third step involves rating each development and content validation study against nine criteria for good content validity (**Table 2**).(65) For each of relevance, comprehensiveness and comprehensibility, if ≥85% of an instruments' items fulfil the criteria, the study is deemed to have sufficient (+) evidence; if <85% of items fulfil the criteria, the study is deemed to have insufficient (—) evidence; and if there is inadequate information available or the study quality was inadequate (as identified through risk of bias assessment), the study is deemed to have indeterminate (?) evidence.(65) From this, we will assign an overall content validation score (+, —, ?) which will represent the woman-centricity of PROM and PREM development. Two reviewers will undertake the content validation assessment.

Table 2: Relevance, comprehensiveness and comprehensibility criteria for evaluating the content validity of maternity care instruments(65)

Criteria
Relevance
1. Are the included items relevant to maternity care?

2. Are the included items relevant to women?
3. Are the response options appropriate?
4. Is the recall period appropriate?
Comprehensiveness
5. Are all key concepts included?
Comprehensibility
6. Are the instrument instructions understood by women as intended?
7. Are items and response options understood by women as intended?
8. Are items appropriately worded?
9. Do the response options match the question?

Evaluating the sufficiency of measurement properties

Instruments will next be evaluated according to how well the reported measurement properties (e.g., structural validity) comply with standards of good psychometric properties (**Table 3**).(28) This will indicate whether a PROM or PREM can be considered valid and reliable. Validity is the extent to which an instrument measures what is purports to measure.(67, 68) Reliability is the extent to which participant responses to an instrument can be replicated in unchanging circumstances (consistency).(69) Reliability is also the extent to which an instrument is devoid of measurement error.(70, 71)

Using the COSMIN updated criteria for good measurement properties, psychometric properties will be rated as + (provides sufficient evidence), ─ (provides insufficient evidence), and ? (provides indeterminate evidence) (**Table 3)**. Red text denotes criteria added based on prominence in the literature relative to instrument development and psychometric evaluation. COSMIN's criteria of 'Hypothesis testing for construct validity' has been excluded from Table 3 as the context of maternity care in this study is too broad for the review team to appropriately generate hypotheses suitable for all potential instruments. If a PROM or PREM has several studies reporting on its' psychometric properties, each study will be evaluated individually (according to the reported psychometric properties), and an overall conclusion regarding the quality of the instrument will be provided for each psychometric quality. Any psychometric properties not assessed will be labelled as having 'no

evidence'. Two reviewers will undertake the good psychometric properties assessment. **Table 3:** COSMIN updated criteria for good measurement properties(28)

Measurement property	Rating	Criteria
Structural validity	+	**Classical test theory (CTT)** • Confirmatory factor analysis (CFA): CFI or TLI (or comparable measure) >0.95, **OR** RMSEA <0.06, **OR** SMSR <0.08; **AND/OR** • Exploratory factor analysis (EFA) or Principal Components Analysis (PCA): KMO ≥0.70, **AND** Significant Bartlett's Test of Sphericity (p<0.05), **AND** dimensional (total) variance explained ≥50% or dimensional (total) variance explained <50% but justified by the authors(72, 73) **Item Response Theory (IRT)/ Rasch:** • No violation of unidimensionality: CFI or TLI (or comparable measure) >0.95, **OR** RMSEA <0.06, **OR** SMSR <0.08; **AND** • No violation of local independence: residual correlations among the items after controlling for the dominant factor <0.20 OR Q_3 fit statistics <0.37; **AND** • No violation of monotonicity: adequate looking graphs, OR item scalability >0.30; **AND** • Adequate model fit – IRT: χ^2 >0.01; Rasch: infit and outfit mean squares between ≥0.50 and ≤1.50, OR z-standardised values between >-2 and <2
	?	**CTT** • Not all information for + reported **IRT/ Rasch** • Model fit not reported
	–	**CTT** • Criteria for + not achieved **IRT/ Rasch** • Criteria for + not achieved
Internal consistency	+	Evidence of sufficient structural validity achieved (+ or ? for 'Structural validity'); **AND** Cronbach's alpha(s) ≥0.70 for each unidimensional scale or subscale
	?	Evidence of sufficient structural validity not achieved
	–	Evidence of sufficient structural validity achieved (+ or ? for 'Structural validity'); **AND** Cronbach's alpha(s) <0.70 for each unidimensional scale or subscale
Reliability	+	ICC or weighted Kappa ≥0.70
	?	ICC or weighted Kappa not reported
	–	ICC or weighted Kappa <0.70
Measurement error	+	SDC or LoA < MIC
	?	MIC not defined
	–	SDC or LoA > MIC
Cross-cultural validity/ measurement invariance	+	No importance differences found between group factors (such as age, gender, language) in multiple group factor analysis; **OR** No important DIF for group factors (McFadden's R^2 <0.02)
	?	No multiple group factor analysis performed; **OR** No DIF analysis performed
	–	Important differences between group factor analysis identified; **OR** Important differences in DIF analysis identified
Criterion validity	+	Correlation with gold standard instrument ≥0.70*; **OR** AUC ≥0.70
	?	Not all information for + reported
	–	Correlation with gold standard instrument <0.70*; **OR** AUC <0.70

Responsiveness	+	AUC ≥0.70
	?	AUC not reported
	–	AUC <0.70

*Correlation with a gold standard will only occur if a short-form instrument is being compared against its long-form counterpart; CTT = Classical Test Theory; CFA = Confirmatory Factor Analysis; CFI = Comparative Fit Index;/ TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SMSR = Standardised Root Mean Residuals; EFA = Exploratory Factor Analysis; CA = Principal Components Analysis; KMO = Kaiser-Meyer-Olkin; IRT = Item Response Theory; ICC = Intraclass Correlation Coefficient; SDC = Smallest Detectable Change; LoA = Limits of Agreement; MIC = Minimally Important Change; DIF = Differential Item Functioning; Area Under the Curve

Summarise and grade the quality of evidence

By summarising and grading the evidence available for an individual instrument, we can provide an overall conclusion as to the quality of that instrument. Thus, this will involve combining the results of each instruments' risk of bias, content validity, and psychometric property assessments into a single metric of 'high', 'moderate', 'low', or 'very low' evidence using the Grading of Recommendations Assessment, Development, and Evaluation (GRAE) approach.(28) If the results across multiple studies pertaining to a single instrument are consistent, then results will be quantitatively pooled and a GRADE score will be reported. If results are inconsistent, they will not be pooled, no GRADE score will be reported, and areas of inconsistency will be discussed (e.g., if an instrument demonstrates differing levels of quality depending on the country in which it is used).

Step 8: Describe the interpretability and feasibility of instrument implementation

Interpretability is the extent to which meaning can be derived from participant responses to an instrument or changes in responses.(74) This may include distinct patterns of responses amongst subgroups of the population, trends in responses over time, and floor or ceiling effects. For the purposes of this review, we will extract and describe the following features of PROM and PREM interpretability: (i) distribution of responses in the study population and relevant subgroups; (ii) proportion of missing data for items; (iii) methods of handling missing data; (iv) evidence of floor and ceiling effects; and (v) minimally important changes (MIC) or minimally important differences (MID) in responses. Interpretability, whilst not considered a measurement property in and of itself, is

important for understanding the real-world application and biases associated with implementing PROMs and PREMs.

Feasibility refers to the ease and convenience with which a PROM or PREM can be implemented and administered in a real-world context.(28) For the purposes of this review, we will extract and describe the following features of PROM and PREM feasibility: (i) available modes of administration; (ii) length of the instrument; (iii) estimated completion time; (iv) level of readability; (v) ease of response calculation; (vi) copyright; (vii) cost of using an instrument; (viii) equipment required for instrument administration; (ix) availability of instrument for application in different settings and languages; and (x) approvals required before instrument use. For the development of a maternity PROMs and PREMs database, this information will be critical for informing the real-world implementation of maternity PROMs and PREMs across health services and systems.

Patient and public involvement statement

The research team comprises members of Maternity Choices Australia, a national consumer advocacy organisation committed to the advancement of best-practice maternity care.(75) These women are consumer representatives and have been involved in the conceptualisation of the research and protocol development, recognising the importance of operationalising woman-centred care, and ensuring that maternity services are consumer informed. Importantly, they will aid the development of the Maternity PROMs and PREMs database, supporting its usability by a range of maternity care stakeholders. They will also help disseminate the Maternity PROMs and PREMs database through formal and informal engagement with key collaborative parties.

ETHICS AND DISSEMINATION

Ethical permission for this research is not required as the review will only use information from previously published research. The findings of this research will be submitted for publication in an international, peer-reviewed journal. Abstracts will also be submitted for national and international conference presentations.

Maternity PROMs and PREMs database

We intend for the maternity PROMs and PREMs database to be freely accessible online, and useful to all individuals involved in maternity health services and systems performance measurement, and value-based maternity care. The design of the database will be consumer informed to ensure that it is easy to understand, and provides information relevant to a range of maternity care stakeholders. The psychometric results (structural validity, internal consistency, reliability, measurement error, cross-cultural validity/ measurement invariance, criterion validity and responsiveness) for each instrument will be summarised according to whether criteria were met (+), indeterminate (?), or not met (—) when all evidence for a specific instrument in considered collectively. The woman-centricity of instrument development will be similarly summarised according to the COSMIN criteria for good content validity. In addition, the database will summarise descriptive information for each instrument (e.g., number of items, domains captured, country of development); summarise information regarding each instruments' feasibility of use (e.g., copyright and reuse considerations, available modes of administration, costs etc.); and provide links to the studies describing instruments. For PROMs or PREMs not freely available, we will also provide the appropriate contact information for the instruments' original author or licensing agent.

We anticipate that the database will be updated annually. A member of the research team will re-run the search strategies (updating search terms as needed) and undertake the processes described in this protocol. This will support the identification of new instruments or additional evidence of PROM and PREM psychometric evaluation over time, ensuring that the database is up-to-date and aligns with advancements in PROM and PREM methodologies and results.

REFERENCES

1. Leap N. Woman-centred or women-centred care: Does it matter? British Journal of Midwifery. 2009;17(1):12-6.

2. Snowden A, Martin C, Jomeen J, Hollins Martin C. Concurrent analysis of choice and control in childbirth. BMC Pregnancy Childbirth. 2011;11:40.

3. Brady S, Lee N, Gibbons K, Bogossian F. Woman-centred care: An integrative review of the empirical literature. Int J Nurs Stud. 2019;94:107-19.

4. Fontein-Kuipers Y, de Groot R, van Staa A. Woman-centered care 2.0: Bringing the concept into focus. Eur J Midwifery. 2018;2(May).

5. Akileswaran CP, Hutchison MS. Making Room at the Table for Obstetrics, Midwifery, and a Culture of Normalcy Within Maternity Care. Obstet Gynecol. 2016;128(1):176-80.

6. Maffi I, Gouilhers S. Conceiving of risk in childbirth: Obstetric discourses, medical management and cultural expectations in Switzerland and Jordan. Health Risk Soc. 2019;21(3-4):185-206.

7. Porter ME, Teisberg EO. Redefining health care: creating value-based competition on results. Brighton, MA: Harvard business press; 2006.

8. Perfetto EM. ISPOR's Initiative on US Value Assessment Frameworks: A Missed Opportunity for ISPOR and Patients. Value Health. 2018;21(2):169-70.

9. Perfetto EM, Oehrlein EM, Boutin M, Reid S, Gascho E. Value to Whom? The Patient Voice in the Value Discussion. Value Health. 2017;20(2):286-91.

10. De Jonge A, Downe S, Page L, Devane D, Lindgren H, Klinkert J, et al. Value based maternal and newborn care requires alignment of adequate resources with high value activities. BMC Pregnancy Childbirth. 2019;19(1):428.

11. Fawsitt CG, Bourke J, Greene RA, McElroy B, Krucien N, Murphy R, et al. What do women want? Valuing women's preferences and estimating demand for alternative models of maternity care using a discrete choice experiment. Health Policy. 2017;121(11):1154-60.

12. Fawsitt CG, Bourke J, Lutomski JE, Meaney S, McElroy B, Murphy R, et al. What women want: Exploring pregnant women's preferences for alternative models of maternity care. Health Policy. 2017;121(1):66-74.

13. Renfrew MJ, McFadden A, Bastos MH, Campbell J, Channon AA, Cheung NF, et al. Midwifery and quality care: Findings from a new evidence-informed framework for maternal and newborn care. Lancet. 2014;384(9948):1129-45.

14. Organisation for Economic Co-operation and Development. Measuring what matters: the patient-reported indicator surveys. Paris: OECD; 2019.

15. Terwee CB, Prinsen CAC, Chiarotto A, de Vet HCW, Bouter LM, Alonso J, et al. COSMIN methodology for assessing the content validity and PROMs: User manual. Amsterdam: COSMIN; 2018.

16. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. Bja Educ. 2017;17(4):137-44.

17. Black N, Burke L, Forrest CB, Sieberer UH, Ahmed S, Valderas JM, et al. Patient-reported outcomes: pathways to better health, better services, and better societies. Qual Life Res. 2016;25(5):1103-12.

18. Williams K, Sansoni J, Morris D, Grootemaat P, Thompson C. Patient-reported outcome measures: Literature review. Sydney: ACSQHC; 2016 November 2016.

19. Centre for Health Services and Policy Research. What are PROMs? Vancouver: The University of British Columbia; [cited 2021 June]. Available from: https://patientreportedoutcomes.ca/what-are-pros/proms/.

20. Sultan P, Sadana N, Sharawi N, Blake L, El-Boghdadly K, Falvo A, et al. Evaluation of Domains of Patient-Reported Outcome Measures for Recovery After Childbirth: A Scoping and Systematic Review. JAMA Netw Open. 2020;3(5):e205540.

21. Sultan P, Sharawi N, Blake L, Ando K, Sultan E, Aghaeepour N, et al. Use of Patient-Reported Outcome Measures to Assess Outpatient Postpartum Recovery: A Systematic Review. JAMA Netw Open. 2021;4(5):e2111600.

22. Sultan P, Ando K, Sultan E, Hawkins J, Blake L, Barwick F, et al. A systematic review of patient-reported outcome measures used to assess sleep in postpartum women using Consensus Based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines. Sleep. 2021;44(10).

23. Sultan P, Ando K, Sultan E, Hawkins JE, Chitneni A, Sharawi N, et al. A systematic review of patient-reported outcome measures to assess postpartum pain using Consensus Based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines. Br J Anaesth. 2021;127(2):264-74.

24. Sharawi N, Klima L, Shah R, Blake L, Carvalho B, Sultan P. Evaluation of patient-reported outcome measures of functional recovery following caesarean section: a systematic review using the consensus-based standards for the selection of health measurement instruments (COSMIN) checklist. Anaesthesia. 2019;74(11):1439-55.

25. Dickinson F, McCauley M, Smith H, van den Broek N. Patient reported outcome measures for use in pregnancy and childbirth: a systematic review. Bmc Pregnancy Childb. 2019;19.

26. Bull C, Byrnes J, Hettiarachchi R, Downes M. A systematic review of the validity and reliability of patient-reported experience measures. Health Serv Res. 2019;54(5):1023-35.

27. Beecher C, Devane D, White M, Greene R, Dowling M. Women's experiences of their maternity care: A principle- based concept analysis. Women Birth. 2020;33(5):419-25.

28. Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, de Vet HCW, et al. COSMIN methodology for systematic reivews of patient-reported outcome measures (PROMs). Amsterdam: COSMIN; 2018.

29. Talmon A, Ginzburg K. "Who does this body belong to?" The development and psychometric evaluation of the Body Experience during Pregnancy Scale. Body Image. 2018;26:19-28.

30. Lundsberg LS, Schwarz EB, Vilardo NA, Yonkers KA, Gariepy AM. Clinical Validation of PROMIS Global Short Form in Pregnancy. Appl Res Qual Life. 2018;13(1):89-103.

31. Bobevski I, Rowe H, Clarke DM, McKenzie DP, Fisher J. Postnatal demoralisation among women admitted to a hospital mother-baby unit: validation of a psychometric measure. Archives of Women's Mental Health. 2015;18(6):817-27.

32. Gartner FR, de Miranda E, Rijnders ME, Freeman LM, Middeldorp JM, Bloemenkamp KW, et al. Good reliability and validity for a new utility instrument measuring the birth experience, the Labor and Delivery Index. J Clin Epidemiol. 2015;68(10):1184-94.

33. Symon A, Downe S, Finlayson KW, Knapp R, Diggle P, team St. The feasibility and acceptability of using the Mother-Generated Index (MGI) as a Patient Reported Outcome Measure in a randomised controlled trial of maternity care. BMC Med Res Methodol. 2015;15:100.

34. Stuge B, Jenssen HK, Grotle M. The Pelvic Girdle Questionnaire: Responsiveness and Minimal Important Change in Women With Pregnancy-Related Pelvic Girdle Pain, Low Back Pain, or Both. Phys Ther. 2017;97(11):1103-13.

35. Cox JL, Holden JM, Sagovsky R. Detection of Postnatal Depression: Development of the 10-item Edinburgh Postnatal Depression Scale. British Journal of Psychiatry. 1987;150(6):782-6.

36. Schaap T, Bloemenkamp K, Deneux-Tharaux C, Knight M, Langhoff-Roos J, Sullivan E, et al. Defining definitions: a Delphi study to develop a core outcome set for conditions of severe maternal morbidity. BJOG: An International Journal of Obstetrics & Gynaecology. 2019;126(3):394-401.

37. Briscoe KE, Haas DM. Developing a Core Outcome Set for Cesarean Delivery Maternal Infectious Morbidity Outcomes. Am J Perinatol. 2020;37(4):436-52.

38. Martin CH, Fleming V. The birth satisfaction scale. Int J Health Care Qual Assur. 2011;24(2):124-35.

39. Vardavaki Z, Hollins Martin CJ, Martin CR. Construct and content validity of the Greek version of the Birth Satisfaction Scale (G-BSS). Journal of Reproductive and Infant Psychology. 2015;33(5):488-503.

40. Göncü Serhatlıoğlu S, Karahan N, Hollins Martin CJ, Martin CR. Construct and content validity of the Turkish Birth Satisfaction Scale - Revised (T-BSS-R). J Reprod Infant Psychol. 2018;36(3):235-45.

41. Sudhinaraset M, Landrian A, Afulani PA, Phillips B, Diamond-Smith N, Cotter S. Development and validation of a person-centered abortion scale: the experiences of care in private facilities in Kenya. BMC Womens Health. 2020;20(1):208.

42. Bull C. Patient satisfaction and patient experience are not interchangeable concepts. Int J Qual Health Care. 2021;33(1).

43. Jonsson A, Orwelius L, Dahlstrom U, Kristenson M. Evaluation of the usefulness of EQ-5D as a patient-reported outcome measure using the Paretian classification of health change among patients with chronic heart failure. J Patient Rep Outcomes. 2020;4(1):50.

44. Benson T, Williams DH, Potts HW. Performance of EQ-5D, howRu and Oxford hip & knee scores in assessing the outcome of hip and knee replacements. BMC Health Serv Res. 2016;16(1):512.

45. NHS Digital. Patient Reported Outcome Measures (PROMs) in England - 2011-2012, Special topic, EQ-5D™ Index analysis UK: NHS; 2014 [updated 12 October 2018; cited 2021 June]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms/patient-reported-outcome-measures-proms-in-england-2011-2012-special-topic-eq-5d-index-analysis.

46. Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Serv Insights. 2013;6:61-8.

47. Feeny DH, Eckstrom E, Whitlock EP, Perdue LA. Patient-Reported Outcomes, Health-Related Quality of Life, and Function: An Overview of Measurement Properties. A Primer for Systematic Reviewers on the Measurement of Functional Status and Health-Related Quality of Life in Older Adults. AHRQ Methods for Effective Health Care. Rockville (MD): AHRQ; 2013.

48. Whitehurst DG, Noonan VK, Dvorak MF, Bryan S. A review of preference-based health-related quality of life questionnaires in spinal cord injury research. Spinal Cord. 2012;50(9):646-54.

49. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. Br Med Bull. 2010;96:5-21.

50. Jain SS, DeFroda SF, Paxton ES, Green A. Patient-Reported Outcome Measures and Health-Related Quality-of-Life Scores of Patients Undergoing Anatomic Total Shoulder Arthroplasty. J Bone Joint Surg Am. 2020;102(13):1186-93.

51. Williams B, Coyle J, Healy D. The meaning of patient satisfaction: an explanation of high reported levels. Soc Sci Med. 1998;47(9):1351-9.

52. Noyes RW, Levy MI, Chase CL, Udry JR. Expectation fulfillment as a measure of patient satisfaction. American Journal of Obstetrics and Gynecology. 1974;118(6):809-14.

53. Williams B. Patient satisfaction: a valid concept? Soc Sci Med. 1994;38(4):509-16.

54. Centers for Medicare & Medicaid Services. HCAHPS Hospital Consumer Assessment of Healthcare Providers and Systems Baltimore, MD: CMS; 2021 [updated 2 February 2021; cited 2021 June]. Available from: https://hcahpsonline.org/en/survey-instruments/.

55. U.S. Centers for Medicare & Medicaid Services. Emergency Department CAHPS (ED CAHPS) Baltimore, MD: CMS.gov; 2020 [updated 11 May 2020; cited 2021 June]. Available from: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/CAHPS/ED.

56. Care Quality Commission. Adult Inpatient Survey 2020 UK: NHS; 2020 [cited 2021 June]. Available from: https://nhssurveys.org/surveys/survey/02-adults-inpatients/.

57. Grogan S, Conner M, Willits D, Norman P. Development of a questionnaire to measure patients' satisfaction with general practitioners' services. Br J Gen Pract. 1995;45(399):525-9.

58. Kinnersley P, Stott N, Peters T, Harvey I, Hackett P. A comparison of methods for measuring patient satisfaction with consultations in primary care. Fam Pract. 1996;13(1):41-51.

59. Hojat M, Louis DZ, Maxwell K, Markham FW, Wender RC, Gonnella JS. A brief instrument to measure patients' overall satisfaction with primary care physicians. Fam Med. 2011;43(6):412-7.

60. Lavrakas PJ. Acquiescence Response Bias. In: Lavrakas PJ, editor. Encyclopedia of Survey Research Methods. Thousand Oaks, CA: Sage Publications; 2008.

61. Dykema J, Schaeffe NC, Garbarski D, Assad N, Blixt S. Towards a reconsideration of the use of agree-disagree questions in measuring subjective evaluations. Res Social Adm Pharm. 2021.

62. Terwee CB, Jansma EP, Riphagen, II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009;18(8):1115-23.

63. Covidence.org. Covidence Melbourne: Covidence.org; 2021 [cited 2021 September]. Available from: https://www.covidence.org/.

64. COSMIN. COSMIN database of systematic reviews of outcome measurement instruments Amsterdam: University Library Vrje Universiteit Amsterdam; 2019 [updated Septmeber 2019; cited 2021 September]. Available from: https://database.cosmin.nl/.

65. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159-70.

66. DeVellis RF. Scale development: Theory and applications. 4th Ed. ed. Bickman L, Rog DJ, editors. Thousand Oaks, California: SAGE Publications, Inc.; 2017. 262 p.

67. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. Health Econ. 1999;8(1):41-51.

68. Babbie ER. The Practice of Social Research. Boston, MA, UNITED STATES: Cengage Learning US; 2015.

69. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155-63.

70. Streiner DL, Norman GR. Health Measurement Scales: A practical guide to their development and use. Oxford Scholarship Online: Oxford Scholarship Online; 2008.

71. American Psychological Association. Reliability Washington, DC: APA; 2020 [cited July 2021]. Available from: https://dictionary.apa.org/reliability.

72. Yong AG, Pearce S. A Beginner's guide to factor analysis: Focusing on Exploratory Factor Analysis. Tutor Quant Methods. 2013;9(2):79-94.

73. Gartner FR, Bomhof-Roordink H, Smith IP, Scholl I, Stiggelbout AM, Pieterse AH. The quality of instruments to assess the process of shared decision making: A systematic review. PLoS One. 2018;13(2):e0191747.

74. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.

75. Maternity Choices Australia. About Brisbane, Australia2021 [cited 2021 October]. Available from: https://www.maternitychoices.org/about.

Funding statement: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests: None declared.

Author contributions: CB contributed to the conception of the research and wrote the protocol manuscript. HT and LC provided important intellectual content in revisions of the protocol manuscript. AR and SC provided important consumer representation and content relative to woman-centred care. EC contributed to the conception of the research and provided important intellectual content in revisions of the protocol manuscript. All authors approved of the version submitted for publication and agree to be accountable for all aspects of the research.

Data sharing: Not applicable.

SUPPLEMENTARY FILE 1 – MEDLINE (OVID) ELECTRONIC DATABASE SEARCH STRATEGIES

Maternity PROMs (07-10-2021)

n	Search terms	Results
1	((Obstetrics/) OR (Prenatal Care/) OR (Maternal Health Services/) OR (Pregnancy/) OR (Delivery, Obstetric/) OR (Postnatal Care/) OR (Pregnant women/) OR (Parturition/) OR (Labor, Obstetric/) OR (Delivery, obstetric/) OR (Obstetric?.ab,ti.) OR (Matern$.ab,ti.) (Prenatal.ab,ti.) OR (Antepartum.ab,ti.) OR (Perinatal.ab,ti.) OR (Intrapartum.ab,ti.) OR (Postnatal.ab,ti.) OR (Postpartum.ab,ti.) OR (Pregnan$.ab,ti.) OR (Childbirth?.ab,ti.) OR (Birth$.ab,ti.))	1,356,332
2	((Patient Reported Outcome Measures/) OR ((patient reported outcome?).ab,ti.) OR ((patient-reported outcome?).ab,ti.) OR (("patient-centred outcome?" or "patient centred outcome?").ti,ab.) OR (("patient-centered outcome?" or "patient centered outcome?").ti,ab.))	22,352
3	(((instrumentation or methods).fs.) OR (("Validation Studies" or "Comparative Study").pt.) OR (Psychometrics/) OR (psychometr$.ti,ab.) OR ((clinimetr$ or clinometr$).tw.) OR (Outcome Assessment, Health Care/) OR (outcome measure$.tw.) OR (Observer Variation/) OR (observer variation.ti,ab.) OR (Health Status Indicators/) OR (Reproducibility of Results/) OR (reproducib$.ti,ab.) OR (Discriminant Analysis/) OR (outcome assessment$.ti,ab.) OR ((reliab$ or unreliab$ or valid$ or coefficient$ or homogeneity or homogeneous or internal consistency).ti,ab.) OR ((cronbach$ and alpha$).ti,ab.) OR ((item and (correlation$ or selection$ or reduction$)).ti,ab.) OR ((agreement or precision or imprecision or "precise values" or test$retest).ti,ab.) OR (test retest.ti,ab.) OR ((reliab$ and (test or retest)).ti,ab.) OR ((stability or inter$rater or intra$rater or inter$tester or intra$tester or inter$observer or intra$observer or inter$technician or intra$technician or inter$examiner or intra$examiner or inter$assay or intra$assay or inter$individual or intra$individual or inter$participant or intra$participant or kappa$ or repeatab$).ti,ab.) OR (((replicab$ or repeated) and (measure$ or finding$ or result$ or test$)).ti,ab.) OR ((generaliza$ or generalisa$ or concordance).ti,ab.) OR (intraclass correlation$.ti,ab.) OR ((discriminative or "known group" or "factor analys$" or dimension$ or subscale$).ti,ab.) OR (multitrait scaling analys$.ti,ab.) OR (("item discriminant" or "interscale correlation$" or error$ or "individual variability").ti,ab.) OR ((variability and (analysis or values)).ti,ab.) OR ((uncertainty and (measurement or measuring)).ti,ab.) OR (("standard error of measurement" or sensitiv$ or responsive$).ti,ab.) OR (((minimal$ or clinical$) and (important or significant or detectable) and (change or difference)).ti,ab.) OR ((small$ and (real or detectable) and (change or difference)).ti,ab.) OR (("meaningful change" or "ceiling effect$" or "floor effect$" or "Item response model$" or IRT or Rasch or "Differential item functioning" or DIF or "computer adaptive testing" or "item bank" or "cross$cultural equivalence").ti,ab.))	8,796,309
4	(("Surveys and Questionnaires"/) OR (questionnaire?.ab,ti.) OR (tool.ab,ti.) OR (scale?.ab,ti.) OR (measure$.ab,ti.) OR (instrument?.ab,ti.))	4,331,661
5	(2 AND 3 AND 4)	14,185
6	(1 AND 5) Limited to: English language; Publication year 2010- current	**111**

? = Wildcard that stands for 0 or 1 replacement character within or at the end of a word; $ = Wildcard that searches for variations of a word of unlimited character length; / = Medical Subject Heading (MeSH) term; ti = Title search; ab = Abstract search

Maternity PREMs (07-10-2021)

n	Search terms	Results
1	((Obstetrics/) OR (Prenatal Care/) OR (Maternal Health Services/) OR (Pregnancy/) OR (Delivery, Obstetric/) OR (Postnatal Care/) OR (Pregnant women/) OR (Parturition/) OR (Labor, Obstetric/) OR (Delivery, obstetric/) OR (Obstetric?.ab,ti.) OR (Matern$.ab,ti.) (Prenatal.ab,ti.) OR (Antepartum.ab,ti.) OR (Perinatal.ab,ti.) OR (Intrapartum.ab,ti.) OR (Postnatal.ab,ti.) OR (Postpartum.ab,ti.) OR (Pregnan$.ab,ti.) OR (Childbirth?.ab,ti.) OR (Birth$.ab,ti.))	1,356,332
2	((“patient reported experience?”.ab,ti.) OR (“patient-reported experience?”.ab,ti.) OR ((wom?n? experience?).ab,ti.) OR ((woman$ experience?).ab,ti.))	9,041
3	(((instrumentation or methods).fs.) OR (("Validation Studies" or "Comparative Study").pt.) OR (Psychometrics/) OR (psychometr$.ti,ab.) OR ((clinimetr$ or clinometr$).tw.) OR (Outcome Assessment, Health Care/) OR (outcome measure$.tw.) OR (Observer Variation/) OR (observer variation.ti,ab.) OR (Health Status Indicators/) OR (Reproducibility of Results/) OR (reproducib$.ti,ab.) OR (Discriminant Analysis/) OR (outcome assessment$.ti,ab.) OR ((reliab$ or unreliab$ or valid$ or coefficient$ or homogeneity or homogeneous or internal consistency).ti,ab.) OR ((cronbach$ and alpha$).ti,ab.) OR ((item and (correlation$ or selection$ or reduction$)).ti,ab.) OR ((agreement or precision or imprecision or "precise values" or test$retest).ti,ab.) OR (test retest.ti,ab.) OR ((reliab$ and (test or retest)).ti,ab.) OR ((stability or inter$rater or intra$rater or inter$tester or intra$tester or inter$observer or intra$observer or inter$technician or intra$technician or inter$examiner or intra$examiner or inter$assay or intra$assay or inter$individual or intra$individual or inter$participant or intra$participant or kappa$ or repeatab$).ti,ab.) OR (((replicab$ or repeated) and (measure$ or finding$ or result$ or test$)).ti,ab.) OR ((generaliza$ or generalisa$ or concordance).ti,ab.) OR (intraclass correlation$.ti,ab.) OR ((discriminative or "known group" or "factor analys$" or dimension$ or subscale$).ti,ab.) OR (multitrait scaling analys$.ti,ab.) OR (("item discriminant" or "interscale correlation$" or error$ or "individual variability").ti,ab.) OR ((variability and (analysis or values)).ti,ab.) OR ((uncertainty and (measurement or measuring)).ti,ab.) OR (("standard error of measurement" or sensitiv$ or responsive$).ti,ab.) OR (((minimal$ or clinical$) and (important or significant or detectable) and (change or difference)).ti,ab.) OR ((small$ and (real or detectable) and (change or difference)).ti,ab.) OR (("meaningful change" or "ceiling effect$" or "floor effect$" or "Item response model$" or IRT or Rasch or "Differential item functioning" or DIF or "computer adaptive testing" or "item bank" or "cross$cultural equivalence").ti,ab.))	8,796,309
4	(("Surveys and Questionnaires"/) OR (questionnaire?.ab,ti.) OR (tool.ab,ti.) OR (scale?.ab,ti.) OR (measure$.ab,ti.) OR (instrument?.ab,ti.))	4,331,661
5	(2 AND 3 AND 4)	1,754
6	(1 AND 5) Limited to: English language; Publication year 2010- current	**496**

? = Wildcard that stands for 0 or 1 replacement character within or at the end of a word; $ = Wildcard that searches for variations of a word of unlimited character length; / = Medical Subject Heading (MeSH) term; ti = Title search; ab = Abstract search

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Checklist item	Page No
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	Identify the report as a protocol of a systematic review	1
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	n/a
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	2
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	1
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	26
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	n/a
Support:			
Sources	5a	Indicate sources of financial or other support for the review	26
Sponsor	5b	Provide name for the review funder and/or sponsor	26
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	26
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	4-5
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	5-6
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	7-8
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	9-10

Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	Supplementary file 1
Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	10
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	10
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	10-15
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	10-15
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	10-15
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	10-15
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	10-15
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)	15
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	n/a
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	n/a
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	n/a
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	n/a

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.